LEUPHANA
UNIVERSITÄT LÜNEBURG

Wer ist leistungsstark? Operationalisierung von Leistungsstärke in der empirischen Bildungsforschung seit dem Jahr 2000

Neuendorf, Claudia; Jansen, Malte; Kuhl, Poldi; Vock, Miriam

Published in:
Zeitschrift fur Padagogische Psychologie

DOI:
10.1024/1010-0652/a000343

Publication date:
2023

Document Version
Verlags-PDF (auch: Version of Record)

Link to publication

Citation for pulished version (APA):
Neuendorf, C., Jansen, M., Kuhl, P., & Vock, M. (2023). Wer ist leistungsstark? Operationalisierung von Leistungsstärke in der empirischen Bildungsforschung seit dem Jahr 2000. *Zeitschrift fur Padagogische Psychologie*, *37*(1-2), 1-19. https://doi.org/10.1024/1010-0652/a000343

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 04. Dez.. 2025

https://econtent.hogrefe.com/doi/pdf/10.1024/1010-0652/a000343 - Sunday, March 19, 2023 12:40:23 AM - Universitätsbibliothek Lüneburg IP Address:141.39.213.142

Übersichtsartikel

Wer ist leistungsstark? Operationalisierung von Leistungsstärke in der empirischen Bildungsforschung seit dem Jahr 2000

Claudia Neuendorf[1], Malte Jansen[2], Poldi Kuhl[3] und Miriam Vock[4]

[1]Hector-Institut für Empirische Bildungsforschung, Eberhard Karls Universität Tübingen, Deutschland
[2]Institut zur Qualitätsentwicklung im Bildungswesen an der Humboldt-Universität zu Berlin, Deutschland
[3]Institut für Bildungswissenschaft, Leuphana Universität Lüneburg, Deutschland
[4]Department Erziehungswissenschaften, Universität Potsdam, Deutschland

Zusammenfassung: Leistungsstarke Kinder und Jugendliche sind in den letzten Jahren zunehmend in den Fokus der Bildungspolitik und der Bildungsforschung gerückt. Allerdings gibt es in der Forschung bislang kein geteiltes Verständnis darüber, was genau unter akademischer Leistungsstärke zu verstehen ist. Die vorliegende Arbeit gibt einen systematischen Überblick darüber, wie Forschende, die seit dem Jahr 2000 die Gruppe der leistungsstarken Schülerinnen und Schüler erforschten, Leistungsstärke in ihren Studien operationalisiert haben. Dabei wurde insbesondere untersucht, welche Leistungsindikatoren genutzt wurden, ob ein spezifischer Fachbezug hergestellt wurde und welche Cut-off-Werte und Vergleichsmaßstäbe angelegt wurden. Die systematische Datenbanksuche lieferte insgesamt N = 309 Artikel, von denen n = 55 die Einschlusskriterien erfüllten. Die Ergebnisse zeigen, dass eine große Vielfalt in der Operationalisierung von Leistungsstärke vorliegt. Die meistgenutzten Leistungsindikatoren waren Noten und Testwerte, wobei fächerübergreifende und fachspezifische Definitionen beide häufig waren. Die Cut-off-Werte der Studien waren zum Teil schwierig vergleichbar, aber dort, wo ein Populationsbezug hergestellt werden konnte, lag der Median des Populationsanteils Leistungsstarker bei 10 Prozent. Die Studie diskutiert methodische und inhaltliche Rahmenbedingungen, welche sich auf die Operationalisierung von Leistungsstärke und ihre Vergleichbarkeit über Studien hinweg auswirken. Die vorliegende Arbeit schließt mit Empfehlungen zur Operationalisierung von Leistungsstärke.

Schlüsselwörter: Leistungsstarke Schüler_innen, Operationalisierung, Definition, Review, Hochbegabung

Who is a high achiever? Operationalisation of high achievement in empirical educational research since the year 2000

Abstract: In recent years, high-achieving students have received increased attention by researchers, policymakers and practitioners. However, the question of what exactly constitutes high academic achievement is not yet agreed upon by the research community. This paper provides a systematic review of how researchers studying high-achieving students since 2000 have operationalized high academic achievement in their research. In particular, we examined which performance indicators were used, whether achievement was conceived of as subject-specific or general, and which cut-off values and comparison standards were applied. The systematic database search yielded N = 309 articles, n = 55 of which were finally included in the analysis. The present study observed a diversity in the operationalization of performance. The most commonly used indicators of performance were grades and test scores, with cross-domain and subject-specific definitions both being common. Some of the studies' cut-off values were difficult to compare, but in instances where a population norm could be derived, the median proportion of high achievers was 10 percent. The study discusses that constraints on generalizability and comparability between different studies on high achievers can arise due to methodological differences. This paper concludes with recommendations for the operationalization of high achievement.

Keywords: High-achieving students, operationalization, definition, review, giftedness

Einleitung

Leistungsstarke Schülerinnen und Schüler sind in den letzten Jahren zunehmend in den Fokus von Bildungspolitik, Bildungsforschung und Bildungspraxis geraten. So wurde im Jahr 2015 die *Strategie zur Förderung leistungsstarker und potenziell leistungsstarker Schülerinnen und Schüler* von der Kultusministerkonferenz [KMK] verabschiedet. Bund

© 2022 The Author(s) Distributed as a Hogrefe OpenMind article under the license CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)

https://econtent.hogrefe.com/doi/pdf/10.1024/1010-0652/a000343 - Sunday, March 19, 2023 12:40:23 AM - Universitätsbibliothek Lüneburg IP Address:141.39.213.142

und Länder haben im Jahr 2016 zudem die gemeinsame *Initiative zur Förderung leistungsstarker und potenziell besonders leistungsfähiger Schülerinnen und Schüler* beschlossen. Ziel dieser Initiative, die derzeit unter dem Namen „LemaS“ („Leistung macht Schule“) läuft, ist es, in den Schulen Ansätze zu entwickeln, um leistungsstarke und potenziell leistungsstarke Kinder und Jugendliche künftig angemessener zu fördern und so auch ihren Anteil an der Schülerschaft zu erhöhen. Diese Entwicklung soll durch Begleitforschung unterstützt werden (Bundesministerium für Bildung und Forschung [BMBF] & KMK, 2016). Forschung, die sich mit der Gruppe der leistungsstarken Schülerinnen und Schüler beschäftigen will, unterliegt dabei der Notwendigkeit, sowohl einen theoretischen Begriff von Leistung zugrunde zu legen als auch eine Operationalisierung von Leistungsstärke zu finden. Um feststellen zu können, ob entwickelte Fördermaßnahmen ihre Ziele erreichen, muss somit eine begründete und nachvollziehbare Festlegung dazu getroffen werden, welche Kinder und Jugendlichen als leistungsstark kategorisiert werden. Auch bei der Rezeption wissenschaftlicher Literatur ist es wichtig, die jeweils zugrunde gelegte Definition von Leistungsstärke zu kennen, um einordnen zu können, ob sich die berichteten Ergebnisse zwischen Studien sinnvoll aufeinander beziehen lassen. Während sowohl hochbegabte als auch leistungsschwache Schülerinnen und Schüler schon viel untersucht wurden, ist die Auseinandersetzung mit leistungsstarken Schülerinnen und Schülern ein vergleichsweise junges, sich gerade erst entwickelndes Forschungsfeld (Köller & Baumert, 2017) und es finden sich bislang noch keine Konventionen zu ihrer Definition innerhalb der Wissenschaftscommunity.

Im vorliegenden Beitrag wird daher ein systematischer Überblick über deutschsprachige und internationale empirische Studien gegeben, die seit dem Jahr 2000 die Gruppe der leistungsstarken Schülerinnen und Schüler untersucht haben. Im Mittelpunkt der Übersichtsarbeit steht die Frage, welche Operationalisierung von Leistungsstärke Forschende in ihren Untersuchungen gewählt haben.

Das Verhältnis von Leistungsstärke und Begabung

Obwohl ein intuitives Verständnis dafür existiert, wann jemand exzellente Leistungen erbringt, liegt keine einheitliche wissenschaftliche Definition dazu vor, was unter Schulleistung zu verstehen ist (Brühwiler & Helmke, 2018). Forschungsarbeiten, die sich mit Kindern und Jugendlichen beschäftigen, welche herausragende Leistungen im akademischen Bereich zeigen, beziehen sich häufig auf die Tradition der Begabungsforschung. Diese entstand zeitgleich mit der Entwicklung der ersten Intelligenztests und bezog sich entsprechend auf die Untersuchung und Beschreibung von Personen, die Höchstleistungen in kognitiven Grundfähigkeiten erbringen, wie sie typischerweise von Intelligenztests gemessen werden (Lubinski, 2016; Terman, 1954). In der Begabungsforschung war lange Zeit die Setzung weithin akzeptiert, die besten zwei Prozent einer Population bzw. Personen, die wenigstens zwei Standardabweichungen über dem Populationsmittelwert liegen, als hochbegabt zu definieren (Amelang & Schmidt-Atzert, 2006; Rost & Buch, 2018). Kognitive Begabung wird dabei als zu einem größeren Anteil genetisch determiniert angesehen (Galton, 1892; Terman, 1926, 1954). Konzeptionell davon abgegrenzt wurde häufig Leistungsstärke (also das Erbringen sehr guter Leistungen in der Schule bzw. in curriculumsnahen, bereichsspezifischen Tests), die sich aus hoher kognitiver Begabung ergeben kann, es aber nicht zwangsläufig tun muss (Hany, 2012). Die Gegenüberstellung von begabten und leistungsstarken Schülerinnen und Schülern erfolgte beispielsweise prominent im Marburger Hochbegabtenprojekt (Rost, 2009), in dem hochbegabte und hochleistende Schülerinnen und Schüler in ihrer Entwicklung miteinander verglichen wurden. In dieser Studie wurden aus der Hochleistungsstichprobe 12 Prozent der Jugendlichen ausgeschlossen, da sie gleichzeitig einen IQ von über 130 hatten, also das Hochbegabungskriterium erfüllten.

Aktuelle Begabungsmodelle, wie das Megamodell von Subotnik, Olszewski-Kubilius und Worrell (2011), schlagen ein verändertes Konzept von Hochbegabung vor, indem sie von der einseitigen Fokussierung auf Intelligenz Abstand nehmen. Stattdessen wird der Prozess der Talententwicklung beschrieben, welcher sich in einer Entwicklung von allgemeineren kognitiven Grundfähigkeiten und Prädispositionen hin zu immer spezialisierterem Wissen und Können vollzieht, um schlussendlich Expertise in einer Domäne hervorzubringen. Leistungsstärke ist aus dieser Perspektive sowohl Resultat früherer Investitionen in ein Talent als auch Prädiktor für eine erfolgreiche Weiterentwicklung des Talents bei entsprechender Förderung. Leistungsstärke wird damit als Maß für Begabung in einem bestimmten Entwicklungsabschnitt und einer bestimmten fachlichen Domäne begriffen. Für die Erforschung von Talententwicklung, wie sie beispielsweise im TAD-Framework (Talent Development in Achievement Domains Framework, Preckel et al., 2020) angeregt wird, ist die Bestimmung von fachbezogener Leistungsstärke damit ähnlich zentral wie die Messung breit angelegter kognitiver Begabung. Um Ergebnisse von Talententwicklungsprozessen zu bewerten und Vorhersagen über den weiteren Erfolg in einer bestimmten Talentdomäne zu machen, stellt sich daher die Frage, wie Leistungsstärke in einem Fach operationalisiert werden soll.

© 2022 The Author(s) Distributed as a Hogrefe OpenMind article under the license CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)

https://econtent.hogrefe.com/doi/pdf/10.1024/1010-0652/a000343 - Sunday, March 19, 2023 12:40:23 AM - Universitätsbibliothek Lüneburg IP Address:141.39.213.142

Diese neuen Entwicklungen in der Begabungsforschung anerkennend, wird aus forschungspraktischen Gründen in der vorliegenden Arbeit dennoch eine Abgrenzung zwischen Leistungs- und Begabungskonzept getroffen, wie sie beispielsweise in den Moderatorenmodellen von Gagné (1985) oder Heller (2001) angelegt ist. Dabei wird Begabung als weitgehend angeborenes, bereits in jungem Alter vorhandenes, hohes allgemeines kognitives Fähigkeitspotenzial verstanden. Leistungsstärke hingegen wird in seinem engen Verständnis verwendet und auf bereits realisierte hohe Leistungen im akademischen Bereich beschränkt.

Empirische Vorarbeiten

Auch wenn es sich bei kognitiven Fähigkeitstests um prinzipiell kontinuierliche Maße handelt, haben sich historisch in der Begabungsforschung Kriterien für die kategoriale Definition von Hochbegabung (2%, s.o.) entwickelt. Ebenso existieren Kriterien für die Diagnostik von Teilleistungsschwächen und sonderpädagogischen Förderschwerpunkten. Für fachbezogene schulische Leistungsstärke hingegen liegen keine Konventionen vor und bisher existiert noch keine Forschungsarbeit, die eine Übersicht darüber bietet, wie Forschende Leistungsstärke definieren und operationalisieren – d.h. welche Kategorisierungsstrategien sie nutzen, um leistungsstarke Schülerinnen und Schüler in ihrer Forschung zu identifizieren. Ein Beispiel für die Diversität an Definitionen von Leistungsstärke bieten nationale und internationale Large-Scale-Assessments, in denen Aussagen über die Leistungsfähigkeit von Schülerinnen und Schülern beziehungsweise deren Anteile an der Schülerschaft getroffen werden. Trotz der hohen Standardisierung legen auch diese Studien bei ihrer Interpretation der Leistungen teilweise unterschiedliche Operationalisierungen von Spitzenleistungen zugrunde. Zum Teil ist dies ein Resultat unterschiedlicher Kompetenzstufenmodelle bei den unterschiedlichen Assessments: So ist die Kompetenzskala beim IQB-Bildungstrend in fünf inhaltlich definierte Kompetenzbänder unterteilt, während die Kompetenzstufen bei PISA im Lesen von ursprünglich fünf Stufen auf sieben Stufen erweitert wurden, um im oberen und unteren Bereich besser zu differenzieren. Die ursprüngliche fünfte Kompetenzstufe wurde dabei nochmals unterteilt. Diese Anpassung verdeutlicht bereits, dass eine Verschiebung des Interesses an leistungsstarken Schülerinnen und Schülern stattgefunden hat. In der überwiegenden Zahl der Berichte werden zum einen die oberen fünf Prozent und zum anderen der Anteil der Schülerinnen und Schüler auf Kompetenzstufe V (bzw. V und VI bei PISA) als leistungsstark, hochkompetent oder „Spitzengruppe“ bezeichnet (z.B. OECD, 2009). Mitunter gibt es allerdings auch Abweichungen von diesem Vorgehen. In einem Unterkapitel des nationalen Berichts zu PISA 2006 werden die 25 Prozent leistungsstärksten in den Naturwissenschaften „hochkompetent“ genannt (Prenzel, Schütte & Walter, 2007). In Zusatzanalysen zu PISA 2012 werden die besten zehn Prozent in Mathematik bzw. Naturwissenschaften als „Highest-achieving students“ herausgestellt (OECD, 2015). Bei IGLU 2011 (Bos et al., 2012) wurden fachübergreifende Fähigkeitsprofile errechnet und die beiden obersten Profile unter der Bezeichnung „Schülerinnen und Schüler mit hohen Leistungen“ beschrieben.

Systematische Untersuchungen zur Operationalisierung hoher Leistungen bezogen sich bisher meist auf das traditionelle Begabungskonstrukt auf Grundlage kognitiver Fähigkeiten. In dieser Begabungsforschung existieren beispielsweise Arbeiten, die sich mit Identifikationsverfahren für Begabtenförderprogramme beschäftigten (Acar, Sen & Cayirdag, 2016; McBee, 2006; Rothenbusch, Zettler, Voss, Lösch & Trautwein, 2016). Zudem liegen zwei Übersichtsarbeiten vor, die die Operationalisierung von Begabung in Forschungsarbeiten systematisch untersucht haben (Carman, 2013; Ziegler & Raul, 2000). In diesen Übersichtsarbeiten wurde deutlich, dass Intelligenztests und Leistungstests beziehungsweise Schulleistung die häufigsten Methoden der Identifikation Begabter für wissenschaftliche Studien waren, gefolgt von Nominierungen, zum Beispiel durch die Lehrkräfte. Häufig wurden in den Studien mehrere Indikatoren herangezogen, um Begabung festzustellen. Bemerkenswert ist aber auch der Befund, dass sich ein substanzieller Anteil an Studien fand, die keine oder nur ungenügende Angaben zur Operationalisierung von Begabung machten.

Herleitung der Forschungsfragen

Zur Beantwortung der Fragestellung, wie Leistungsstarke in Studien identifiziert wurden, wird im vorliegenden Beitrag eine systematische Übersichtsarbeit erstellt. Ziel war es, alle Artikel seit dem Jahr 2000, welche leistungsstarke Schülerinnen und Schüler untersuchten, zu identifizieren und deren Operationalisierung des Konstrukts *Leistungsstärke* zu systematisieren.

Den zwei im Folgenden dargestellten Teilfragestellungen wird hierbei nachgegangen.

Forschungsfrage 1: Welche Indikatoren werden zur Operationalisierung von Leistungsstärke herangezogen?

Die erste Teilfrage, die in dieser Arbeit beantwortet werden soll, ist, welche Art von Leistungsmaß zur Identifika-

© 2022 The Author(s) Distributed as a Hogrefe OpenMind article under the license CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)

https://econtent.hogrefe.com/doi/pdf/10.1024/1010-0652/a000343 - Sunday, March 19, 2023 12:40:23 AM - Universitätsbibliothek Lüneburg IP Address:141.39.213.142

tion leistungsstarker Schüler_innen eingesetzt wird (Frage 1.a). Das in der Praxis bedeutsamste Schulleistungsmaß stellen Schulnoten dar. Mittels Schulnoten überführen Lehrkräfte die zu bewertenden Leistungs- und Verhaltensaspekte in Zahlen (Maaz, Baeriswyl & Trautwein, 2013). Sie erfüllen im Bildungssystem verschiedene Funktionen (z. B. pädagogische Funktion, Selektions- und Allokationsfunktion), weshalb es nicht verwunderlich ist, dass neben der schulischen Leistung auch andere Aspekte wie das Arbeitsverhalten und die Motivation in Noten einfließen und nicht nur kriteriale, sondern auch individuelle und soziale Referenzmaßstäbe bei der Vergabe eine Rolle spielen (Hochweber, 2010; Lintorf, 2012a, 2012b; Rüdiger, Jansen & Rjosk, 2021). Daher ist die Vergleichbarkeit von Noten über einzelne Klassen, Schulen und Schulformen hinweg nur bedingt gegeben. Auch die Urteile zur Reliabilität und Validität von Schulnoten fallen ungünstig aus (Ingenkamp & Lissmann, 2008; Trapmann, Hell, Weigand & Schuler, 2007). Den Schulnoten werden häufig objektive und standardisierte Leistungs- und Kompetenztests gegenübergestellt. Hierzu gehören zum Beispiel Kompetenztests aus nationalen und internationalen Large-Scale-Assessments, die dem Monitoring und der Evaluation des Bildungssystems dienen (z. B. Stanat, Schipolowski, Mahler, Weirich & Henschel, 2019; Reiss, Weis & Klieme, 2019; Schwippert et al., 2020), aber auch weitere normierte und standardisierte Tests, die sich auch zur Individualdiagnostik eignen (z. B. ELFE II, Lenhard, Lenhard & Schneider, 2020, DEMAT 2+, Krajewski, Dix & Schneider, 2020, BASIS-MATH, Moser Opitz, Stöckli, Grob, Nührenbörger & Reusser, 2019). Es ist anzunehmen, dass die Operationalisierung von Leistungsstärke durch Noten und Kompetenz- und Leistungstests einen großen Anteil der Studien charakterisieren wird. Bei den Überblicksarbeiten aus dem Bereich der Begabungsforschung waren Nominierungen durch Lehrkräfte ebenfalls zu einem kleineren Anteil vertreten.

Ob häufiger Einzelindikatoren oder häufiger – wie bei Ziegler und Raul (2000) – eine Kombination mehrerer Indikatortypen (z. B. Noten und Testleistungen) eingesetzt werden, ist eine weitere Frage, die untersucht werden soll (Frage 1.b). McBee und Makel (2019) wiesen darauf hin, dass die Kombination mehrerer Indikatoren zu einem Hochbegabungs-Index, je nach Korrelation untereinander und genutzter Kombinationsregel, einen bedeutsamen Einfluss auf den Anteil Hochbegabter hat. Gleichzeitig wird häufig empfohlen, in der Diagnostik multiple Instrumente einzusetzen, um eine höhere Reliabilität und Validität zu erreichen (Amelang & Schmidt-Atzert, 2006; Schmidt & Hunter, 1998). Gleiches gilt für die Operationalisierung von Leistungsstärke.

Weitergehend soll die Domänenspezifität des Konstrukts Leistungsstärke untersucht werden. Im Gegensatz zur traditionellen Begabungsforschung, in der Intelligenz als fachunabhängige kognitive Leistungsfähigkeit im Fokus steht, muss Leistung im Sinne von Performanz sich erst in einem oder mehreren Fächern manifestieren, um festgestellt zu werden. Dabei ist es nicht ungewöhnlich, dass Leistungen einer Schülerin oder eines Schülers in verschiedenen Fächern unterschiedlich ausfallen. So erreichte beispielsweise im Bildungstrend 2016 etwa ein Drittel der leistungsstarken Grundschülerinnen und Grundschüler (30 %, bzw. 7 % der Gesamtpopulation) sowohl in Mathematik als auch in Deutsch die höchste Kompetenzstufe in beiden Fächern. 27 Prozent waren hingegen nur in Mathematik und 43 Prozent nur im Fach Deutsch leistungsstark (Neuendorf, Kuhl & Jansen, 2017). Eine weitere Studie, welche leistungsstarke Jugendliche anhand von PISA-Daten der Jahre 2000 und 2003 untersuchte, stellte fest, dass lediglich etwa drei Prozent der Population in allen drei Inhaltsdomänen (Lesen, Mathematik und Naturwissenschaft) zum leistungsstärksten Zehntel gehörten (Zimmer, Brunner, Lüdtke, Prenzel & Baumert, 2007). Der fachliche Bezug wird auch im TAD-Framework (Preckel et al., 2020) hervorgehoben, welches die Erforschung und Förderung von Leistungsentwicklung in spezifischen Leistungsdomänen zum Ziel hat.

Gleichzeitig kann Hochleistung auch als Gegenstück zur Hochbegabung verstanden und damit ebenfalls domänenübergreifend konzeptualisiert werden (Rost, 2009). Zum Beispiel untersuchten Zimmer et al. (2007) „vielseitig hochkompetente“ Schülerinnen und Schüler, die in allen Kompetenzbereichen (Mathematik, Lesen und Naturwissenschaften) zu den Leistungsstärksten gehörten. Mit Verweis auf die hohen Korrelationen zwischen Leistungen in verschiedenen Fächern definierten Köller und Baumert (2017) für die BERLIN-Studie Leistungsstärke aufgrund eines fächerübergreifenden Faktorscores. Damit stellt sich für die vorliegende Übersichtsarbeit die Frage, ob Forschende Leistungsstärke eher fachspezifisch oder fächerübergreifend definieren (Frage 1.c).

Forschungsfrage 2: Welche Cut-off-Werte werden zur Bestimmung von Leistungsstärke herangezogen?

Carman (2013) stellte in seiner Untersuchung fest, dass ein Großteil der Studien im Bereich der Begabungsforschung – obwohl für die kognitive Hochbegabung mit dem 2 %-Kriterium zumindest eine teilweise geteilte Konvention vorlag – entweder keine Cut-off-Werte oder keine Verteilung der Testwerte in einer Normstichprobe angegeben hatte und auch das genaue Testinstrument selten berichtet wurde. Da damit häufig kaum nachzuvollziehen ist, wer als begabt kategorisiert wurde, sei eine Interpretierbarkeit

© 2022 The Author(s) Distributed as a Hogrefe OpenMind article under the license CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)

https://econtent.hogrefe.com/doi/pdf/10.1024/1010-0652/a000343 - Sunday, March 19, 2023 12:40:23 AM - Universitätsbibliothek Lüneburg IP Address:141.39.213.142

der Ergebnisse und deren Übertragbarkeit auf andere Stichproben gefährdet. Diesem zentralen Kritikpunkt soll auch in der vorliegenden Untersuchung in Bezug auf Leistungsstärke nachgegangen werden. Dazu werden die Studien dahingehend kodiert, ob und welche Cut-off-Werte für die jeweiligen Indikatoren berichtet werden. In diesem Zusammenhang besonders relevant ist weiterhin die Frage, welche Art von Bezugsnorm verwendet wird. Bei einer kriterialen Bezugsnorm etwa würden absolute Notenstufen, Testwerte oder Kompetenzstufen verwendet. Bei einer sozialen Referenznorm wird die relative Position in der Leistungsverteilung genutzt. Dazu wird ebenfalls untersucht, ob die Referenz für den jeweiligen Cut-off-Wert eine Populationsnorm, die Stichprobe der jeweiligen Untersuchung oder sogar die Leistungsverteilung innerhalb von Subgruppen (z. B. auf Klassenebene) ist.

Methode

Unser Vorgehen orientiert sich an den bei Gough (2007) beschriebenen Schritten eines Systematic Reviews. *Phase 1* dient nach Gough (2007) der Entwicklung der Fragestellungen. In *Phase 2: Ein- und Ausschlusskriterien* wurden Studien in die Analyse eingeschlossen, welche die folgenden Kriterien erfüllten: 1) Es musste sich um einen empirischen Artikel in einer wissenschaftlichen Fachzeitschrift handeln. Diese Einschränkung erfolgte, um sicherzustellen, dass alle eingeschlossenen Beiträge ein Peer-Review-Verfahren durchlaufen haben. 2) Das Veröffentlichungsdatum musste zwischen 2000 und 2020 liegen. Die Beschränkung auf diesen Zeitraum erfolgte, um einen relativ aktuellen Überblick zu erhalten. 3) Die Stichprobe musste Schülerinnen und Schüler im Primar- oder Sekundarbereich umfassen. 4) Der Artikel musste mindestens eine Operationalisierung der Zielgruppe mit überdurchschnittlichen Schulleistungen (z. B. Leistungsstarke, high achievers, high performers) enthalten (z. B. durch die Stichprobenauswahl mit einem Fokus auf Leistungsstarke oder durch Bildung von Leistungsgruppen innerhalb einer größeren Stichprobe). 5) Es musste sich primär um Leistungen oder Kompetenzen in den Hauptfächern handeln ([fremd-]sprachliche, mathematische, naturwissenschaftliche Kompetenzen). Entsprechend wurden Studien, die sich eindeutig beispielsweise auf Hochleistungen im sportlichen oder musikalischen Bereich bezogen, ausgeschlossen. 6) Da die Operationalisierung von schulischer Leistungsstärke im engeren Sinne im Fokus stand, wurden Studien ausgeschlossen, die sich konzeptionell ausschließlich auf Begabte (z. B. durchgängige Nutzung der Begriffe „gifted", „high cognitive ability" statt „achievement" oder „performance") bezogen oder die ausschließlich Intelligenztestmaße, aber keine Maße zur Erfassung schulischer Leistungen heranzogen (z. B. Studien, die auf Basis der Study of Mathematically Precocious Youth, SMPY, entstanden sind, z. B. Lubinski & Benbow, 2006).

Phase 3: Recherchestrategie. Zunächst wurden deutschsprachige und internationale Datenbanken im Bereich Erziehungswissenschaft und Psychologie nach relevanten Studien durchsucht. Die Datenbanken waren im Einzelnen *PsycArticles*, *Psyndex Literature & AV Media* und *ERIC*. Da die unterschiedlichen Datenbanken unterschiedliche Thesauri für Schlagwörter und unterschiedliche Filtermöglichkeiten bieten, unterschieden sich die Suchbegriffe je nach Datenbank (s. Tabelle A1 im Anhang). Zusätzlich wurden weitere Artikel aufgenommen, die zum Beispiel von den Studien, die in den Datenbanken gefunden worden waren, zitiert wurden und die Einschlusskriterien dieses Systematic Review erfüllten. Die Suchen lieferten insgesamt 309 Ergebnisse.

Phase 4: Screening. Die Studien wurden in einem mehrschrittigen Verfahren gescreent und bei mangelnder Passung aussortiert. In einem ersten Schritt des Screenings wurde lediglich der Titel des Beitrags betrachtet. Auf diese Weise wurden Studien ausgeschlossen, bei denen bereits im Titel deutlich wurde, dass es sich nicht um Schülerinnen und Schüler im allgemeinbildenden Schulsystem handelte, dass es nicht um Schulleistungen in den Hauptfächern ging oder dass keine empirische Originalarbeit vorlag. In einem zweiten Schritt wurde dann das Abstract und in einem dritten Schritt der Methodenteil des Artikels untersucht. 144 Artikel, welche die oben genannten Ein- und Ausschlusskriterien erfüllten, verblieben nach dem Screening von Abstract und Methodenteil in der weiteren Analyse.

Das Screening wurde für einen zufällig ausgewählten Anteil von 10 % der Studien zusätzlich von einem zweiten Bewerter durchgeführt. Die Beobachterübereinstimmung betrug 83 % (Cohen's $\kappa = .66$). Unterschiede waren vor allem darin begründet, dass der Zweitbewerter das Kriterium 5 (Fachbezug) strenger auslegte und bei Notendurchschnitten ohne genauere Benennung der Fächer einen Fallausschluss empfahl, während die Erstautorin solche Studien in die Bewertung einschloss und nur solche Studien ausschloss, in denen es eindeutig nicht um Leistungen in Hauptfächern ging. Im Zweifelsfall wurden diese Studien zunächst eingeschlossen, da die Beurteilung der Qualität der Darstellung ein Teil der inhaltlichen Analyse werden sollte und auch das Wissen darüber, wie viele Studien eine unvollständige Darstellung der Operationalisierung von Leistungsstärke aufwiesen, von Interesse war.

Schließlich wurden die 144 verbleibenden Artikel danach bewertet, inwiefern sie ihren inhaltlichen Fokus tatsächlich auf leistungsstarke Schülerinnen und Schüler legten. Ein großer Teil der Studien kategorisierte Schüle-

© 2022 The Author(s) Distributed as a Hogrefe OpenMind article under the license CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)

https://econtent.hogrefe.com/doi/pdf/10.1024/1010-0652/a000343 - Sunday, March 19, 2023 12:40:23 AM - Universitätsbibliothek Lüneburg IP Address:141.39.213.142

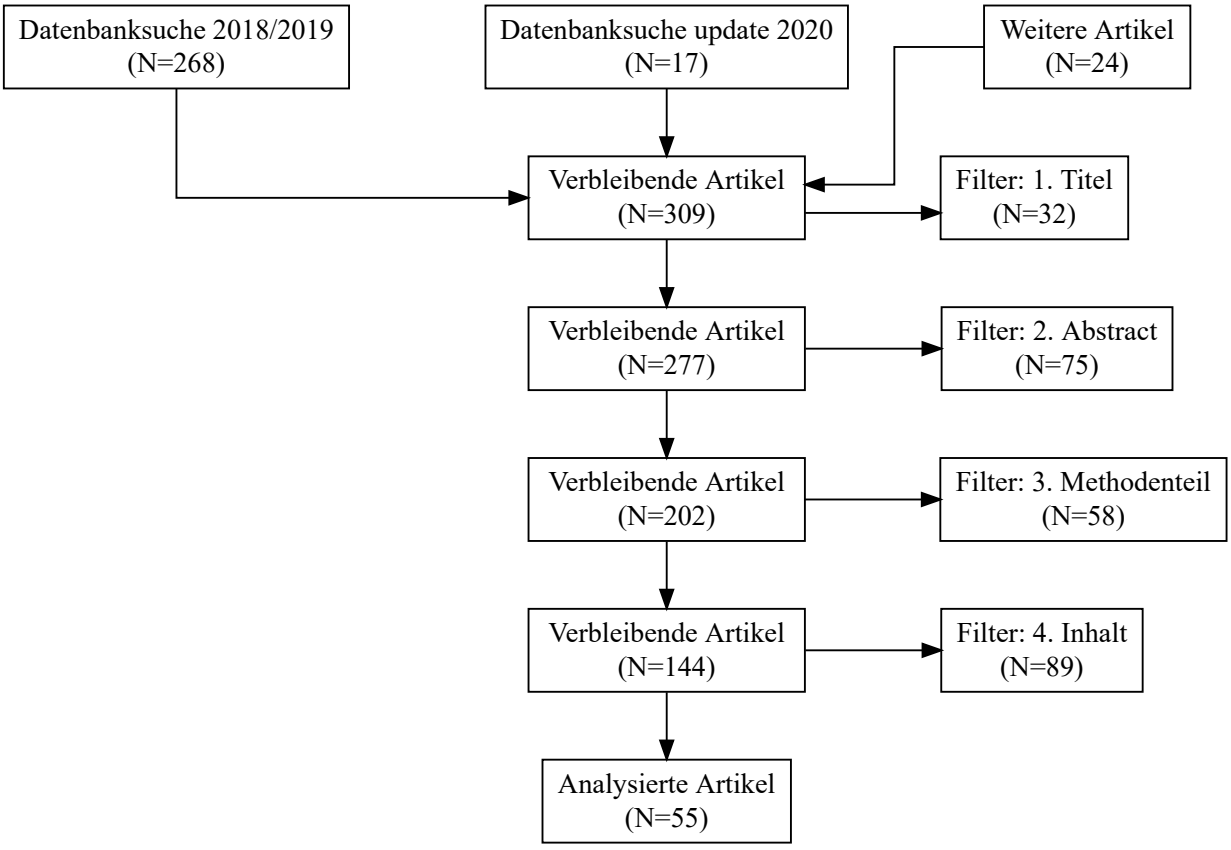

Abbildung 1. Flowchart des Screening-Prozesses. Eine detaillierte Übersicht der Studien, die in den einzelnen Screeningschritten ausgeschlossen wurden, findet sich im ESM 1, der Code zur Reproduktion der Abbildung findet sich unter https://osf.io/jzkv6.

rinnen und Schüler zwar anhand ihrer Leistung und definierte in diesem Rahmen eine leistungsstarke (bzw. leistungsstärkere) Gruppe, allerdings waren leistungsstarke Schülerinnen und Schüler nicht ein Fokus der Untersuchung und die Gruppe wurde häufig sehr breit definiert (z. B. durch einen Median-Split). Vielmehr ging es in diesen Studien um die Robustheit oder Varianz der interessierenden Effekte (z. B. der Wirksamkeit von innerer und äußerer Differenzierung im Unterricht; Trautwein, Köller & Kämmerer, 2002) über das gesamte Leistungsspektrum hinweg. Diese Studien, die im Kern nicht Leistungsstärke untersuchten, wurden ausgeschlossen, so dass sich die Anzahl der Artikel auf die 55 Studien reduzierte, deren inhaltlicher Fokus die Leistungsstärke war (siehe Abbildung 1). Ein Überblick über die Ergebnisse der Literaturrecherche und den Screeningprozess findet sich in den elektronischen Supplementen (ESM) 1 und 2.

Die eingeschlossenen Studien verteilten sich über eine Vielzahl an wissenschaftlichen Zeitschriften. Die häufigsten davon waren: *High Ability Studies* (*N* = 6), *Gifted Child Quarterly* (*N* = 4) und *Journal of Educational Psychology* (*N* = 3). Die Publikationsjahre der Studien zeigen einen Anstieg an Studien ab 2012 mit einem Maximum im Jahr 2015.

Die meisten Studien stammten aus Nordamerika (*N* = 27), gefolgt von Europa (*N* = 16) und Asien (*N* = 7). Drei Studien waren multinational und jeweils eine Studie stammte aus Afrika und Australien. In 10 der 55 Artikel wurden Daten deutscher Schülerinnen und Schüler verwendet.

Phase 6: Datenextraktion. Alle Studien wurden anhand der im Folgenden dargestellten Merkmale kodiert. Diese lassen sich in folgende Bereiche einteilen: 1) Merkmale, die die Operationalisierung beschreiben (Forschungsfragen 1 und 2) und 2) Merkmale, die forschungsmethodische und inhaltliche Aspekte der untersuchten Studien beschreiben. Zur Operationalisierung wurde kodiert, um welche Art von Indikator es sich handelte, auf welche Fächer sich die Leistungsstärke bezog, welcher Cut-off verwendet wurde, welche Bezugsnorm angelegt wurde, welche Stichprobengröße vorlag und welche Anteile Leistungsstarker an Stichprobe und Population sich aus den Angaben der Studie ableiten ließen. Forschungsmethodisch verfolgte der Großteil der Artikel einen quantitativen Ansatz, wobei neben den 47 quantitativ-beobachtenden Studien (72 %) auch 6 (quasi)experimentelle Studien (9 %) vertreten waren. 11 Studien (17 %) waren qualitativ

© 2022 The Author(s) Distributed as a Hogrefe OpenMind article under the license CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)

https://econtent.hogrefe.com/doi/pdf/10.1024/1010-0652/a000343 - Sunday, March 19, 2023 12:40:23 AM - Universitätsbibliothek Lüneburg IP Address:141.39.213.142

Tabelle 1. Verteilung der Studien auf die Altersgruppen bzw. Schulstufen

ISCED-Level	Häufigkeit	Anteil
Level 1: Grundschule	13	20 %
Level 1 – 2: Längsschnitt Grundschule – Sekundarstufe I	4	6 %
Level 2: Sekundarstufe I	25	39 %
Level 1 – 3: Längsschnitt Grundschule – Sekundarstufe II	2	3 %
Level 2 – 3: Längsschnitt Sekundarstufe I – II	7	11 %
Level 3: Sekundarstufe II	14	22 %

Anmerkung: ISCED = International Standard Classification of Education. Die Kategorien wurden gebildet, indem je nach Verfügbarkeit bei den Studien die Angabe zur Altersgruppe, zur Klassenstufe oder zur Schulform auf das entsprechende ISCED-Level transformiert wurde.

Tabelle 2. Indikatoren zur Operationalisierung der Zielgruppe der Leistungsstarken

Indikator	Häufigkeit	Anteil
Test	30	46 %
Noten	22	34 %
Schulkontext	21	32 %
Einschätzungen	2	3 %
Andere	2	3 %
Kombination mehrerer Indikatoren	8	12 %

Anmerkung: N = 65.

und eine verfolgte einen Mixed-Methods-Ansatz (2 %). Die Studien waren zu etwas weniger als der Hälfte (31 Studien, 48 %) querschnittlich ausgerichtet. Die übrigen 23 Studien (35 %) waren Längsschnittstudien, 11 weitere Trendstudien (17 %). Die Stichprobengrößen schwankten sehr stark und lagen zwischen 9 und etwa 500,000 Teilnehmenden, bei einem Median von 616 Schülerinnen und Schülern. Der Großteil der Studien untersuchte Schülerinnen und Schüler in der Sekundarstufe I (39 %). Grundschulen und Sekundarstufe II sowie Studien, die über mehrere Bildungsstufen hinweg reichten, waren in etwa gleich vertreten mit jeweils rund 20 Prozent der Studien (s. Tabelle 1). Eine Übersicht über die Studien inklusive der Kodierungen findet sich im Anhang, Tabelle A2.

Acht der 55 Artikel enthielten eine oder mehrere Teil-Studien, wobei dies entweder eine zweite Stichprobe war, die untersucht wurde, oder unterschiedliche Operationalisierungen für Leistungsstärke an einer Stichprobe angewendet wurden. Diese unterschiedlichen Operationalisierungen oder Stichproben gingen als separate Studien in die Analysen ein, weshalb die Grundgesamtheit an Studien in der Analysestichprobe auf 65 stieg.

Ergebnisse

Forschungsfrage 1: Welche Indikatoren wurden zur Operationalisierung von Leistungsstärke herangezogen?

Die häufigsten Indikatoren von Leistungsstärke, die in den Studien Verwendung fanden, waren (1) Tests, (2) Noten und (3) die Zugehörigkeit zu einer speziellen Schule (z. B. *magnet school*, Schulen speziell für Leistungsstarke mit kompetitivem Auswahlverfahren), einer Schulform (z. B. Gymnasium) oder zu einem bestimmten Kurs für Leistungsstarke (z. B. Honors Kurse oder Advanced Placement Programs). Die letzten drei Kategorien wurden unter der Bezeichnung „Schulkontext" zusammengefasst. In 57 Studien (88 %) wurde lediglich ein einzelner Indikatortyp zugrunde gelegt (Frage 1.b), wobei aber durchaus auch Werte desselben Typs miteinander verrechnet werden konnten (z. B. Notenschnitt über mehrere Fächernoten). In 8 Studien (12 %) wurden hingegen mehrere Indikatortypen kombiniert (s. Tabelle 2). Beispielsweise beschrieben einige der Studien, welche die Zulassung zu einer speziellen Schule zugrunde legten, Zulassungsverfahren mit multiplen Kriterien (z. B. Noten, Motivationsschreiben, Intelligenztests oder Empfehlungsschreiben).

In einer weiteren Analyse wurde der Frage nachgegangen, ob Leistungsstärke in den Studien fachbezogen oder

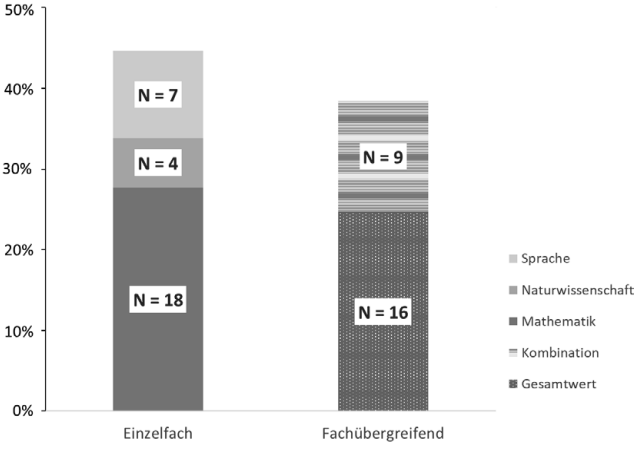

Abbildung 2. Fachbezug der Definitionen von Leistungsstärke. N = 11 Studien machten keine Angaben bzw. nutzten keine Fachleistungen zur Operationalisierung. Mit Gesamtwert sind Studien gemeint, die Notenschnitte bildeten oder einen präexistenten Schnitt (z. B. Abiturgesamtpunktzahl) nutzten. Mit Kombination sind Studien gemeint, die Fachleistungen anders zusammenfassten, z. B. durch Faktorenanalysen, Latente Profilanalysen, und- bzw. oder-Kriterien. Der Code zur Reproduktion der Abbildung findet sich unter *https://osf.io/jzkv6*.

© 2022 The Author(s) Distributed as a Hogrefe OpenMind article under the license CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)

https://econtent.hogrefe.com/doi/pdf/10.1024/1010-0652/a000343 - Sunday, March 19, 2023 12:40:23 AM - Universitätsbibliothek Lüneburg IP Address:141.39.213.142

als fächerübergreifend erfasst wurde (Frage 1.c). Von den untersuchten Studien bezogen sich 45 % in ihrer Definition von Leistungsstärke auf Leistungen in einem Fach, 39 % legten die Leistungen in mehreren Fächern zugrunde. Dabei verwendeten 16 Studien einen Durchschnittswert. In neun Studien wurde Leistungsstärke ebenfalls über mehrere Fächer bestimmt, aber kein Notendurchschnitt verwendet, sondern andere Verfahren wie z. B. Profilanalysen angewandt (s. Abbildung 2).

Forschungsfrage 2: Welche Cut-off-Werte wurden berichtet?

Auch bei einer Verwendung vergleichbarer Indikatoren kann jedoch die Übertragbarkeit von Ergebnissen zwischen verschiedenen Studien fraglich sein, wenn sich die festgelegten Cut-off-Werte zur Bildung von Leistungsgruppen stark unterscheiden. Darüber hinaus kann ein ähnlicher absoluter Cut-off-Wert eine sehr unterschiedliche Bedeutung haben, wenn eine soziale Referenznorm verwendet wird und unterschiedliche Vergleichsgruppen dafür vorliegen. Im Folgenden wird daher die Untersuchung der genutzten Cut-off-Werte gemeinsam mit der Betrachtung der angelegten Vergleichsnormen diskutiert.

Bei quantitativen Indikatoren können Cut-off-Werte zwischen verschiedenen Studien verglichen werden, nachdem sie auf eine gemeinsame Skala transformiert wurden. Bei Testwerten geschieht dies häufig durch Perzentilwerte, die auf der Werteverteilung in der Normstichprobe beruhen. Dies ist allerdings nur bei normierten Tests oder aber in Erhebungen möglich, in denen die Verteilung in der Stichprobe der Verteilung in der Population entspricht. Von den 30 Studien, die Tests eingesetzt haben, haben 27 (90 %) normierte Leistungstests verwendet. Von diesen nutzten 19 Studien Daten aus Large-Scale-Assessment-Studien oder administrative Daten aus staatlichen Lernstandserhebungen. Sieben dieser Studien verwendeten die Angabe der erreichten Kompetenzstufe (zur Entwicklung von Kompetenzstufenmodellen, siehe z. B. Pant, Tiffin-Richards & Köller, 2010) und damit eine kriteriale Bezugsnorm zur Definition von Leistungsstärke. Die übrigen 12 Studien legten eine soziale Bezugsnorm an. Die Cut-off-Werte dieser Studien lagen zwischen 4 und 50 Prozent (*Mdn* = 25 %)

Acht Studien (29.6 %) setzten normierte Tests ein, wie zum Beispiel den Iowa Test of Basic Skills (Hoover, Dunbar & Frisbie, 2001) oder den Woodcock-Johnson Test of Achievement (Woodcock, McGrew & Mather, 2001). Fünf dieser Studien verwendeten die Populationsnormen bei der Beschreibung ihrer Stichprobe der leistungsstarken Schülerinnen und Schüler. Die entsprechenden Cut-off-Werte lagen zwischen 2 und 15 Prozent (*Mdn* = 4 %) und somit deutlich geringer als bei den Studien auf Basis von Large-Scale-Assessments. Drei Studien wählten alternative Identifikationsverfahren: Die eine Studie definierte die besten zwei Prozent auf Klassenebene, die jedoch nicht gleichzeitig über dem 98. Perzentil der Normstichprobe liegen durften, als leistungsstark (für eine Erläuterung s. Rambo-Hernandez & McCoach, 2015) und die anderen beiden Studien nutzten statistische Verfahren, um die Gruppe der Leistungsstarken zu definieren (Latente Profilanalyse: Wang, Eccles & Kenny, 2013, Faktorenanalyse: Robinson, Lanzi, Weinberg, Ramey & Ramey, 2002).

Zusammenfassend lässt sich sagen, dass die Cut-off-Werte bei Tests erheblich streuten. Dort, wo sie auf eine gemeinsame Skala transferierbar waren, nahmen sie Werte zwischen 2 und 50 Prozent an (*Mdn* = 10 %). Bezogen auf die Stichproben der Studien, wurde im Mittel etwa ein Siebtel der Schülerinnen und Schüler mithilfe von Testwerten als leistungsstark definiert (*M* = 15 %, *SD* = 12 %).

Von den 22 Studien (34 %), die Noten nutzten, um die Gruppe der Leistungsstarken zu identifizieren, nutzten 16 Studien einen konkreten Notenwert und damit im Prinzip einen kriterialen Cut-off-Score, 3 Studien gaben einen Anteil leistungsstarker Schülerinnen und Schüler vor (2 %: Forgasz & Hill, 2013; Vock, Köller & Nagy, 2013, bzw. 10 %: Sontag & Stoeger, 2015). In einer Studie (Mourgues, Hein, Tan, Diffley III & Grigorenko, 2016) wurde eine latente Klassenanalyse durchgeführt, um über das höchstgelegene Leistungsprofil Leistungsstärke zu bestimmen. Zwei weitere Studien machten keine Angaben zum Cut-off-Wert. Die eingesetzten Notenskalen waren notwendigerweise divers und schwer vergleichbar, da die Stichproben aus unterschiedlichen Ländern stammten (s. Tabelle A3 im Anhang). Forschende versuchten, diese Schwierigkeit zu umgehen, indem sie das jeweilige Bewertungssystem möglichst genau beschrieben, indem sie eine Umrechnung in den in den USA geläufigen GPA (Grade Point Average) vornahmen oder indem der Anteil an Schülerinnen und Schülern angegeben wurde, die diese Noten typischerweise erhalten (z. B. Salmela & Uusiautti, 2015). Von den 14 Studien (22 %), die ausschließlich Noten zur Einteilung der Leistungsstarken verwendeten, bestand die Stichprobe in fünf Studien komplett aus Leistungsstarken (die Auswahl erfolgte also direkt bei der Stichprobenziehung). Bei weiteren acht Studien lagen die Stichprobenanteile Leistungsstarker zwischen 3 und 50 Prozent (*M* = 29 %). Studien, die neben Noten weitere Kriterien verwendeten, legten einen geringeren Cut-off-Wert bei den Noten an.

Bei der Operationalisierung über den Schulkontext ist in der Regel Hintergrundwissen über diesen vonnöten, um die Exklusivität der Gruppe beurteilen zu können. In manchen Studien (*N* = 12) wurden Kinder und Jugendliche an speziellen Schulen für Leistungsstarke untersucht. Hier ist

© 2022 The Author(s) Distributed as a Hogrefe OpenMind article under the license CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)

https://econtent.hogrefe.com/doi/pdf/10.1024/1010-0652/a000343 - Sunday, March 19, 2023 12:40:23 AM - Universitätsbibliothek Lüneburg IP Address:141.39.213.142

eine Vergleichbarkeit zu anderen Studien schwierig, da häufig idiosynkratische Fallbeschreibungen die Datengrundlage bildeten. Teils wurde angegeben, wie umfangreich das Auswahlverfahren einer Schule ist oder welche Leistungen zu erbringen sind, um zugelassen zu werden. In manchen Studien werden auch Auswahlquoten angegeben. In lediglich einem kleinen Teil der Studien geht es dabei allerdings um Fragestellungen, die ganz explizit eine konkrete Schule oder Schulform betreffen.

Diskussion

Vollständigkeit der Angaben zur Operationalisierung

Das vorliegende Review fragte danach, wie in den Studien der letzten 20 Jahre Leistungsstärke operationalisiert wurde. Ausgewählt wurden insbesondere die Studien, deren erklärtes Ziel es war, leistungsstarke Schülerinnen und Schüler in schulischen Hauptdomänen zu untersuchen. Die Studien zeichneten sich durch eine große Diversität in ihren methodischen Merkmalen aus. Sie zeigten eine große Streuung in Stichprobenumfang, Forschungs- und Erhebungsdesign und untersuchten verschiedene Altersgruppen im allgemeinbildenden Schulsystem.

Der Detailgrad der Beschreibungen der Leistungsstärke unterschied sich zwischen den Studien sehr stark. Teilweise waren beinahe alle relevanten Informationen bereits im Abstract enthalten, teilweise im Methodenteil vorhanden, teilweise musste jedoch im gesamten Artikel nach den relevanten Informationen gesucht werden. Ziegler und Raul (2000) fanden bei 11% der Studien keine Angaben zur Identifikation Begabter, während in den hier untersuchten Studien nur vereinzelt relevante Informationen gänzlich fehlten. Zwei Studien machten keine Angaben zum gewählten Cut-off in der Stichprobe, drei Studien machten keine Angabe dazu, welche Fächer zur Bestimmung der Leistungsstärke betrachtet wurden, bei sieben Studien fehlten Angaben zur Stichprobengröße und in fünf Studien wurden keine Angaben zum Anteil Leistungsstarker an der Stichprobe gemacht. Während einige der fehlenden Angaben lediglich die Interpretierbarkeit des Begriffs von Leistungsstärke erschweren, der zugrunde lag (z. B. was bedeutet eine bestimmte Note? Welchem Anteil an der Population entspricht ein bestimmter Cut-off?), ist das komplette Fehlen von Angaben, die die Identifikation nachvollziehbar machen, eine Ausnahme und trat bei lediglich zwei Studien auf. Bei diesen Studien wurde lediglich berichtet, dass Leistungsstarke aufgrund ihrer Schulleistungen identifiziert wurden (Adeloun et al., 2015; Mioduser & Betzner, 2008).

Diskussion zentraler Ergebnisse

Forschungsfrage 1: Welche Indikatoren werden zur Operationalisierung von Leistungsstärke herangezogen?

Im vorliegenden Review zu leistungsstarken Schülerinnen und Schülern wurden als Indikatoren für Leistungsstärke am häufigsten Tests und Noten herangezogen, die auch in der Übersichtsarbeit von Carman (2013) zur Klassifikation von Begabten eine wichtige Rolle gespielt hatten. Die Vorauswahl Leistungsstarker durch das Bildungssystem, beispielsweise durch Selektionsprozesse in segregierten Bildungssystemen, war in der vorliegenden Übersichtsarbeit beinahe ebenso häufig wie die Identifikation Leistungsstarker anhand von Noten. Bei Ziegler und Raul (2000) nutzte die Mehrheit der Studien multiple Kriterien zur Identifikation Begabter; im vorliegenden Beitrag fand eine Kombination mehrerer Indikatoren hingegen lediglich bei einem kleinen Teil der Studien statt. Zusammenfassend lässt sich schlussfolgern, dass leistungsstarke Schülerinnen und Schüler eine Gruppe darstellen, die sich zwar zunächst konzeptionell von begabten (im Sinne von hochintelligenten) Schülerinnen und Schülern unterscheiden lässt: Bei Begabung liegt die Betonung stärker auf einem Potenzial, welches noch nicht ausgeschöpft sein muss, während bei Leistungsstärke die Performanz und damit bereits gezeigte Leistungen im Mittelpunkt stehen. Bei deren Identifikation in der Forschungspraxis gibt es aber durchaus große Überschneidungen. So wäre es basierend auf den Ergebnissen der Reviews in vielen Fällen schwierig, allein von der Operationalisierung einer beforschten Gruppe darauf zu schließen, ob Forschende hier begabte oder leistungsstarke Schülerinnen und Schüler untersuchen wollten. Die Entwicklungen hin zu einer stärkeren Integration von Begabung und Leistung im Rahmen neuerer Talententwicklungsmodelle (Preckel et al., 2020; Subotnik et al., 2011) scheinen vor diesem Hintergrund folgerichtig.

Eine weitere Hauptfrage in diesem Zusammenhang war die Domänenspezifität von Leistungsstärke. Es fanden sich sowohl Studien, die sich auf Leistungen in einem einzelnen Fach bezogen als auch Studien, die mehrere Fächer in ihre Definition von Leistungsstärke einbezogen. Das Spannungsfeld zwischen der fachspezifischen Manifestation von Leistung und dem fächerübergreifenden Konzept der „Einserschülerin“ bzw. des „Einserschülers“ äußert sich auch darin, dass einige Studien die Analysen für jedes Fach separat wiederholten, aber auch die Bildung von Leistungsprofilen über mehrere Fächer hinweg keine Seltenheit war. Neben den theoretischen Differenzen, die mit diesen unterschiedlichen Ansätzen verbunden sind und die Vergleichbarkeit erschweren, gehen die Ansätze auch mit unterschiedlichen Vor- und Nachteilen einher. So ver-

© 2022 The Author(s) Distributed as a Hogrefe OpenMind article under the license CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)

https://econtent.hogrefe.com/doi/pdf/10.1024/1010-0652/a000343 - Sunday, March 19, 2023 12:40:23 AM - Universitätsbibliothek Lüneburg IP Address:141.39.213.142

bessert die Aggregation von Einzelnoten zwar die Reliabilität der Messung, Rückschlüsse auf Stärken und Schwächen von Schülerinnen und Schülern gehen aber verloren. Profilanalysen ermöglichen es zwar, diese Stärken und Schwächen zu berücksichtigen, können aber ebenfalls eine geringe Reliabilität aufweisen, wenn etwa wenige Variablen in die Berechnung der Profile eingehen und wenn die einzelnen Merkmale hoch korreliert sind (Conger & Lipshitz, 1973).

Zusätzliche Auswertungen zeigten, dass an Grundschulen häufiger Notenschnitte und in der Sekundarstufe I häufiger Einzelfächer oder Fächerprofile zur Klassifikation Leistungsstarker genutzt wurden. Legt man aktuelle Begabungsmodelle zugrunde, so passen beide Beobachtungen – sowohl die Überschneidung der beiden Konstrukte Leistungsstärke und Begabung, als auch die fachspezifische und fachübergreifende Betrachtung von Leistungsstärke – in den Talententwicklungsansatz aktueller Begabungsmodelle (Preckel et al., 2020; Subotnik et al., 2011), welche Begabung als Entwicklung von einer breit angelegten (kognitiven) Potenz hin zur Ausbildung immer spezialisierterer Kompetenzen und Fähigkeiten begreifen.

Forschungsfrage 2: Welche Cut-off-Werte werden verwendet?

Bei der Definition eines angemessenen Cut-off-Wertes zur Abgrenzung der leistungsstarken Schülerinnen und Schüler auf dem Leistungsspektrum zeigte sich am deutlichsten, was Forschende unter Leistungsstärke verstehen. Dabei muss unterschieden werden zwischen der Festlegung eines Anteils Leistungsstarker an der Stichprobe und einem Cut-off-Wert, der einen Rückschluss auf die Anteile in der Population und damit eine Vergleichbarkeit über Studien hinweg zulässt. Dass dies unterschiedliche Aspekte sind, zeigt sich beispielsweise daran, dass viele Studien deutliche Abweichungen zwischen dem Stichprobenanteil Leistungsstarker und dem angenommenen Populationsanteil Leistungsstarker aufwiesen. Um diesen Populationsparameter feststellen zu können, müssen detaillierte Informationen zu den eingesetzten Tests und der Stichprobe vorliegen. Eine Übertragbarkeit des Stichprobenanteils Leistungsstarker auf die Verhältnisse in der Population wird nur dann möglich, wenn Large-Scale-Assessments (ggf. mit Populationsgewichten), Vollerhebungen, eine dezidierte Zufallsauswahl oder normierte Tests vorliegen und eine Populationsnorm als Vergleichsmaßstab zur Verfügung steht. Bei Studien, die Noten als Leistungsindikator nutzen, stellte sich dieses Problem in besonderem Ausmaß, da Noten, wie eingangs erwähnt, kaum zwischen verschiedenen Klassen, noch weniger aber häufig zwischen unterschiedlichen Ländern vergleichbar sind.

Bei den Studien, bei denen dieser Rückschluss auf die Population möglich war, waren die Cut-off-Werte generell strenger. Die Hälfte dieser Studien legte einen Cut-off-Wert an, der weniger als 10 % der Population als leistungsstark definierte. Etwa ein Sechstel der Studien definierte zwei Prozent der Population als leistungsstark – es kann vermutet werden, dass Forschende sich mit diesem strengen Kriterium an der vorherrschenden Definition von Hochbegabung orientierten, die ca. zwei Prozent der Population (IQ-Wert von mindestens 130 Punkten) umfasst (Rost & Buch, 2018). Bei diesen Studien könnte man bereits fragen, inwiefern ein konzeptioneller Unterschied zur Begabung besteht.

Allerdings machte die Kodierung des Cut-off-Kriteriums die größten Schwierigkeiten. Forschende machten häufig unvollständige Angaben, die eine Einschätzung des intendierten (d. h. populationsbezogenen) Cut-off-Wertes verhinderten. Über die Hälfte der Studien ließ keinen Schluss auf das intendierte Cut-off-Kriterium zu. Die Ursache hierfür war meist, dass vorausgelesene Stichproben verwendet und idiosynkratische Indikatoren genutzt wurden, die sehr spezifisch für die jeweilige Stichprobe konstruiert wurden. Teilweise entsteht der Eindruck, teilweise wird explizit gesagt, dass nicht aufgrund inhaltlicher, sondern aufgrund statistischer Überlegungen (Schätzbarkeit mittels parametrischer Verfahren, vergleichbare Gruppengrößen) ein bestimmter Cut-off-Wert gewählt worden war. Teils wurden auch unterschiedliche Cut-off-Werte nebeneinandergestellt, um empirisch den Einfluss der Entscheidung für einen Wert auf die Ergebnisse zu prüfen (Gagné, 2016; Möller & Pohlmann, 2010; Neuendorf et al., 2020; Rambo-Hernandez & McCoach, 2015; Schurtz, Pfost & Artelt, 2014; Zhou, Fan, Wei & Tai, 2017).

Einschränkungen der Generalisierbarkeit

Die Ergebnisse unserer Studie beruhen notwendigerweise auf einem spezifischen Ausschnitt der wissenschaftlichen Literatur. Dieser ist zunächst durch den Untersuchungszeitraum eingeschränkt. Es ist gut möglich, dass die Operationalisierung von Leistungsstärke in der Vergangenheit anders aussah, insbesondere auf Grundlage der sich ändernden Rahmenbedingungen der Forschung (z. B. Verfügbarkeit von Large-Scale-Assessment-Daten, Weiterentwicklung von Theorien und Forschungsparadigmen) und in der Zukunft anders aussehen wird. Ziel unserer Studie war aber die Abbildung des aktuellen Verständnisses des Begriffs Leistungsstärke auch vor dem Hintergrund aktueller bildungspolitischer und fachlicher Diskussionen. Der von uns untersuchte Ausschnitt ist auch stark von den genutzten Fachdatenbanken und Suchbegriffen geprägt. So nutzten wir als einzige deutschsprachige Datenbank Psyndex, entschieden uns aber gegen die Nutzung von FIS-Bildung als deutschsprachige Fachdaten-

© 2022 The Author(s) Distributed as a Hogrefe OpenMind article under the license CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)

https://econtent.hogrefe.com/doi/pdf/10.1024/1010-0652/a000343 - Sunday, March 19, 2023 12:40:23 AM - Universitätsbibliothek Lüneburg IP Address:141.39.213.142

bank im Bereich der Pädagogik. Grund hierfür war in erster Linie die fehlende Möglichkeit bei FIS-Bildung, nach Beiträgen in wissenschaftlichen Fachzeitschriften zu filtern. Weiterhin war der Begriff der Leistungsstärke nicht in der Schlagwortliste von FIS-Bildung enthalten. Wir bemühten uns, durch Variation der Suchbegriffe (top student, high performer, high achiever) eine größere Menge passender Artikel zu identifizieren und entschieden uns schließlich datenbankspezifisch Begriffe zu verwenden, die auf Basis erster Sichtungen der Ergebnisse ein sinnvolles Spektrum an Resultaten ergaben. In einigen Punkten ist die Entscheidung über einen Aus- oder Einschluss einer Studie zu einem gewissen Grad subjektiv – das traf insbesondere auf die Frage zu, welche Studien einen hinreichenden Fokus auf leistungsstarke Schülerinnen und Schüler aufwiesen. Wir bemühten uns daher, möglichst umfassend den Rechercheprozess zu dokumentieren, um unser Vorgehen nachvollziehbar und prüfbar zu machen (s. Tabelle A1 im Anhang, elektronische Supplemente ESM 1 und ESM 2, sowie weitere Materialien im Open Science Framework unter https://osf.io/jzkv6/).

Schlussfolgerungen

Für die Forschung, die sich mit leistungsstarken Schülerinnen und Schülern befasst, ergeben sich vier zentrale Empfehlungen.

Empfehlung zur theoretischen Herleitung einer Operationalisierung

Das Systematic Review hat verdeutlicht, dass leistungsstarke Schülerinnen und Schüler aus unterschiedlichsten theoretischen Perspektiven und mit unterschiedlichen Zielstellungen untersucht werden. Daraus folgt, dass es *erstens* wichtig ist, zu verdeutlichen, welcher theoretische Begriff von Leistungsstärke bei der Untersuchung im Vordergrund steht. Dieser theoretische Hintergrund sollte die Grundlage der Definition leistungsstarker Schülerinnen und Schüler darstellen und eine Begründung für den Operationalisierungsansatz bilden. Ist beispielsweise der bedeutsame Aspekt für eine spezifische Fragestellung die soziale Konstruktion von Leistungsstärke (zum Beispiel Peer-Interaktionen oder Viktimisierung Leistungsstarker), dann wäre ein niedrig inferentes und klar erkennbares Merkmal, welches Kinder und Jugendliche gegenüber ihren Klassenkameraden offensichtlich als leistungsstark auszeichnet, wie zum Beispiel die Schulnoten, eine soziale Bezugsnorm, und ggf. ein fächerübergreifender Ansatz angemessen. Wenn das Thema der Studie hingegen in der Wirksamkeit bestimmter Unterrichtsmethoden liegt, wäre zu erwarten, dass eher ein fachspezifischer, objektiver und über Klassen hinweg vergleichbarer Testwert als Leistungsindikator genutzt wird. Liegt der Fokus eher auf dem Erreichen von Kompetenzen in einem bestimmten Fach, dann wäre ein Kompetenztest und möglicherweise eine kriteriale Bezugsnorm die vorzuziehende Wahl. Ebenso sollte, wenn das TAD-Modell (Preckel et al., 2020) die theoretische Grundlage einer Untersuchung bildet, die Definition je nach Phase im Talententwicklungsprozess domänenübergreifend (frühe Phasen) oder fachspezifisch sein (fortgeschrittene Phasen). Hierbei ist auch zu beachten, dass bestimmte inhaltliche Fragestellungen an Untersuchungsdesigns gebunden sind, welche wiederum die Art der Operationalisierung von Leistungsstärke mit bedingen. So benötigen beispielsweise Fragestellungen, die sich mit mikrosoziologischen Prozessen befassen, häufiger einen qualitativen Zugang, der wiederum spezielle Rahmenbedingungen (z. B. reichhaltige Kontextinformationen, Möglichkeit zum Einbezug multipler Perspektiven / Intersubjektivität) für die Definition und Operationalisierung von Leistungsstärke schafft. Bei anderen Fragestellungen (wie zum Beispiel zu gesellschaftlichen Disparitäten) sind repräsentative, teils länderübergreifende Stichproben essenziell, welche sich insbesondere in Large-Scale-Assessment Studien finden lassen. In den Large-Scale-Assessments liegen fachspezifische Testwerte vor, die auch im hohen Leistungsbereich gut differenzieren. Daher und aufgrund der großen Stichprobenumfänge ist die Möglichkeit gegeben, auch kleinere Gruppen leistungsstarker Schülerinnen und Schüler auf belastbare Weise zu identifizieren.

Obgleich eine gemeinsame Definition des Begriffs bisher nicht existiert und eine Vereinheitlichung über alle Studien hinweg kaum realistisch ist, wäre ein geteiltes Verständnis des Konstrukts Leistungsstärke zumindest innerhalb von Forschungsfeldern ein Fortschritt und würde dafür sorgen, dass Forscherinnen und Forscher besser auf vorangegangenen Forschungsarbeiten aufbauen können. Besonders die Forschungsfelder, die sich mit der Entwicklung und Förderung von Kindern und Jugendlichen mit herausragenden akademischen Leistungen befassen, wie zum Beispiel die neuere Begabungsforschung, würden von einer einheitlichen Definition von Leistungsstärke profitieren. Angesichts des stärker fachspezifischen Ansatzes der Talentförderung in neueren Begabungsmodellen (Preckel et al., 2020) wäre eine Definition, die eine fachspezifische Operationalisierung erlaubt, aber gleichzeitig Vergleiche zwischen Leistungsstarken unterschiedlicher Talentdomänen ermöglicht, vorzuziehen. Zu diesem Zweck erscheint eine Definition anhand des Populationsanteils (ähnlich wie die Konvention der besten 2 % in der Begabungsforschung) sinnvoll. Wie groß dieser Anteil bemessen sein sollte, ist eine Frage, die noch im Forschungsfeld auszuhandeln ist. Dabei sollte idealerweise die praktische Relevanz der Definition bedacht werden. Bisher

© 2022 The Author(s) Distributed as a Hogrefe OpenMind article under the license CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)

https://econtent.hogrefe.com/doi/pdf/10.1024/1010-0652/a000343 - Sunday, March 19, 2023 12:40:23 AM - Universitätsbibliothek Lüneburg IP Address:141.39.213.142

kursieren einerseits operationale Definitionen, die sich an der Größe der Begabtengruppe orientieren und somit einen sehr kleinen Anteil akademisch leistungsstarker Kinder und Jugendlicher annehmen (Rost, 2009), aber auch Begründungen, die zum Beispiel auf die Kompetenzstufendefinitionen von Large-Scale-Assessments Bezug nehmen und einen größeren Anteil der Schülerinnen und Schüler beschreiben (z. B. Köller & Baumert, 2017; Neuendorf et al., 2020).

Empfehlung zur Darstellung der Operationalisierung im Methodenteil

Unabhängig von der theoretischen Begründung für die Gruppendefinition sollten jedoch *zweitens*, wenn leistungsstarke Kinder und Jugendliche als eigene Gruppe dargestellt werden, die Stichproben- und Instrumentenbeschreibung im Methodenteil so konkret sein, dass Leserinnen und Leser einschätzen können, wie die beschriebene Gruppe definiert worden ist. Falls eine vorausgelesene Stichprobe genutzt wird, sollte erkennbar sein, auf welchen Anteil der Population sich die Aussagen über Leistungsstarke generalisieren lassen (siehe auch Simons, Shoda & Lindsay, 2017). Werden kriteriale Normen genutzt, dann sollte expliziert werden, welche Bedeutung dieses Kriterium besitzt und möglichst auch, welcher Anteil der Population dieses Kriterium typischerweise erreicht.

Empfehlung zur Diskussion der Operationalisierung

Drittens sollte in Forschungsarbeiten diskutiert werden, aus welchen Gründen genau diese Operationalisierung gewählt wurde (z. B. theoretische Überlegungen, statistische Notwendigkeiten). Da Schulleistung ein kontinuierliches Merkmal ist, stellt sich auch die Frage, in welchem Fall eine Gruppeneinteilung überhaupt notwendig ist (z. B. wenn es um die Qualifikation zur Teilnahme an bestimmten Förderangeboten geht) und in welchen Fällen auch das gesamte Leistungsspektrum genutzt werden könnte, um Fragestellungen mit Bezug zu leistungsstarken Schülerinnen und Schülern zu beantworten.

Empfehlung zur Prüfung der Generalisierbarkeit der Befunde

Oftmals lässt sich die Operationalisierung von Leistungsstärke nicht eindeutig aus Theorie und Zielstellung ableiten. Es liegt daher an den Forschenden, die gewählte Operationalisierung zu begründen. Daher wird hier *viertens* und abschließend eine Möglichkeit vorgestellt, die Stabilität von Ergebnissen für verschiedene Operationalisierungen von leistungsstarken Kindern und Jugendlichen zu prüfen. Die Multiversumsanalyse (Steegen, Tuerlinckx, Gelman & Vanpaemel, 2016) ist eine Methode, bei der systematisch verschiedene Operationalisierungen gegenübergestellt werden. Dabei werden Entscheidungen während der Datenaufbereitung systematisch variiert und die Effekte dieser unterschiedlichen Spezifikationen miteinander verglichen. So kann beurteilt werden, ob die Ergebnisse diesen Variationen standhalten. So könnten bei Studien zu leistungsstarken Schülerinnen und Schülern Ergebnisse für verschiedene Konfigurationen nachvollziehbarer Entscheidungen bezüglich der verwendeten Indikatoren, Cut-off-Kriterien, Vergleichsmaßstäbe und Fachbezüge verglichen und die Varianz der Ergebnisse in Abhängigkeit von diesen Parametern abgeschätzt werden. Eine Möglichkeit, die Ergebnisse solcher systematisch permutierten Kombinationen darzustellen und die Effekte visuell zu vergleichen, besteht in der Spezifikationskurve (Simonsohn, Simmons & Nelson, 2020). Diese Darstellung ermöglicht es, herauszufinden, welche Entscheidung den größten Anteil an Variation in den Ergebnissen erklärt (z. B. Neuendorf et al., 2020).

Ausblick

Was Preckel et al. (2020) über die Begabungsforschung schreiben, trifft auch auf die Forschung zu Leistungsstärke im schulischen Kontext zu: Das Feld würde von mehr Koordination zwischen Forschenden und einer stärkeren Verzahnung von Theorie und Praxis profitieren. Leistungsstarke Schülerinnen und Schüler werden in verschiedenen Teilgebieten der Psychologie, wie differenzieller Psychologie, Motivationspsychologie, kognitiver Psychologie, pädagogischer Psychologie und auch Begabungsforschung untersucht. Die vorliegende Übersichtsarbeit benennt mit der transparenten Operationalisierung von Leistungsstärke eine Bedingung für die Beurteilung der Übertragbarkeit von Ergebnissen zwischen den Fachbereichen aber auch zwischen verschiedenen Studien innerhalb von Forschungsgebieten. Dadurch wird die Integration von Wissen aus den verschiedenen Teilbereichen ermöglicht. Darüber hinaus soll die Relevanz der Forschungsergebnisse für Praxisinitiativen, wie dem eingangs benannten LemaS-Projekt, erhöht werden. Auch hier kann die vorliegende Arbeit einen Beitrag leisten, wenn aufbauend auf ihren Resultaten Indikatoren entwickelt werden, anhand derer die Ergebnisse von Förderbemühungen beurteilt werden können.

Elektronische Supplemente (ESM)

Die elektronischen Supplemente sind mit der Online-Version dieses Artikels verfügbar unter https://doi.org/10.1024/1010-0652/a000343

ESM 1. Studienübersicht Screeningprozess. Dieses Dokument enthält die Ergebnisse der Datenbankrecherche gruppiert nach Ausschluss im Screeningprozess (PDF)

© 2022 The Author(s) Distributed as a Hogrefe OpenMind article under the license CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)

https://econtent.hogrefe.com/doi/pdf/10.1024/1010-0652/a000343 - Sunday, March 19, 2023 12:40:23 AM - Universitätsbibliothek Lüneburg IP Address:141.39.213.142

ESM 2. Beurteilerübereinstimmung. Dieses Dokument enthält die Zweitkodierungen des Screeningprozesses (Excel)

Literatur

Acar, S., Sen, S. & Cayirdag, N. (2016). Consistency of the performance and nonperformance methods in gifted identification. *Gifted Child Quarterly, 60*(2), 81 – 101. https://doi.org/10.1177/0016986216634438

Adelodun, G. A. & Asiru, A. B. (2015). Instructional resources as determinants of english language performance of secondary school high-achieving students in Ibadan, Oyo State. *Journal of Education and Practice, 6*(21), 195 – 200.

Allen, W. & Griffin, K. (2006). Mo' Money, Mo' Problems? High-achieving black high school students' experiences with resources, racial climate, and resilience. *Journal of Negro Education, 75*(3), 478 – 494. https://www.jstor.org/stable/40026816

Amelang, M. & Schmidt-Atzert, L. (2006). *Psychologische Diagnostik und Intervention* (Springer-Lehrbuch, 4., vollst. überarb. u. erw. Aufl.). Berlin, Heidelberg: Springer Medizin. https://doi.org/10.1007/3-540-28507-5

Assouline, S. G., Ihrig, L. M. & Mahatmya, D. (2017). Closing the excellence gap: Investigation of an expanded talent search model for student selection into an extracurricular STEM program in rural middle schools. *Gifted Child Quarterly, 61*(3), 250 – 261. https://doi.org/10.1177/0016986217701833

Barron, B. (2000). Problem solving in video-based microworlds: Collaborative and individual outcomes of high-achieving sixth-grade students. *Journal of Educational Psychology, 92*(2), 391 – 398. https://doi.org/10.1037/0022-0663.92.2.391

Bergold, S., Kasper, D., Wendt, H. & Steinmayr, R. (2020). Being bullied at school: the case of high-achieving boys. *Social Psychology of Education, 23*(2), 315 – 338. https://doi.org/10.1007/s11218-019-09539-w

Bergold, S., Wendt, H., Kasper, D. & Steinmayr, R. (2017). Academic competencies. Their interrelatedness and gender differences at their high end. *Journal of Educational Psychology, 109*(3), 439 – 449. https://doi.org/10.1037/edu0000140

Berkowitz, E. & Cicchelli, T. (2004). Metacognitive strategy use in reading of gifted high achieving and gifted underachieving middle school students in New York City. *Education and Urban Society, 37*(1), 37 – 57. https://doi.org/10.1177/0013124504268072

Bos, W., Wendt, H., Ünlü, A., Valtin, R., Euen, B., Kasper, D. et al. (2012). Leistungsprofile von Viertklässlerinnen und Viertklässlern in Deutschland. In W. Bos, I. Tarelli, A. Bremerich-Vos & K. Schwippert (Hrsg.), *IGLU 2011. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 227 – 259). Münster [u. a.]: Waxmann.

Brühwiler, C. & Helmke, A. (2018). Determinanten der Schulleistung. In D. H. Rost, J. R. Sparfeldt & S. Buch (Hrsg.), *Handwörterbuch pädagogische Psychologie* (Beltz Psychologie 2018, 5., überarbeitete und erweiterte Auflage). Weinheim: Beltz.

Bundesministerium für Bildung und Forschung; Kultusministerkonferenz. (2016, 28. September). *Gemeinsame Initiative von Bund und Ländern zur Förderung leistungsstarker und potenziell besonders leistungsfähiger Schülerinnen und Schüler*.

Bütüner, S. Ö. & Filiz, M. (2017). Exploring high-achieving sixth grade students' erroneous answers and misconceptions on the angle concept. *International Journal of Mathematical Education in Science and Technology, 48*(4), 533 – 554. https://doi.org/10.1080/0020739X.2016.1256444

Carman, C. A. (2013). Comparing apples and oranges. Fifteen years of definitions of giftedness in research. *Journal of Advanced Academics, 24*(1), 52 – 70. https://doi.org/10.1177/1932202X12472602https://doi.org/10.1177/1932202X12472602

Carroll, P. E. & Bailey, A. L. (2016). Do decision rules matter? A descriptive study of English language proficiency assessment classifications for English-language learners and native English speakers in fifth grade. *Language Testing, 33*(1), 23 – 52. https://doi.org/10.1177/0265532215576380

Carter, D. J. (2008). Achievement as resistance: The development of a critical race achievement ideology among black achievers. *Harvard Educational Review, 78*(3), 466 – 497.

Carter Andrews, D. J. (2012). Black achievers' experiences with racial spotlighting and ignoring in a predominantly white high school. *Teachers College Record, 114*(10), 1 – 46.

Castejón, A. & Zancajo, A. (2015). Educational differentiation policies and the performance of disadvantaged students across OECD countries. *European Educational Research Journal, 14*(3 – 4),222 – 239.https://doi.org/10.1177/1474904115592489

Cheung, K. C. (2017). The effects of resilience in learning variables on mathematical literacy performance: A study of learning characteristics of the academic resilient and advantaged low achievers in Shanghai, Singapore, Hong Kong, Taiwan and Korea. *Educational Psychology, 37*(8), 965 – 982. https://doi.org/10.1080/01443410.2016.1194372

Clausen, M., Weingarten, J. & Wegner, H. (2013). Unterrichtsqualität an einer besonderen Schule: Videobasierte Evaluation eines Oberstufen-Internats für leistungsstarke und hoch motivierte Schülerinnen und Schüler. *Gruppendynamik und Organisationsberatung, 44*(3), 301 – 321. https://doi.org/10.1007/s11612-013-0218-y

Conger, A. J., & Lipshitz, R. (1973). Measures of reliability for profiles and test batteries. *Psychometrika, 38*(3), 411 – 427.

Crawford, C., Macmillan, L. & Vignoles, A. (2017). When and why do initially high-achieving poor children fall behind? *Oxford Review of Education, 43*(1), 88 – 108. https://doi.org/10.1080/03054985.2016.1240672

Ee, J., Moore, P. J. & Atputhasamy, L. (2003). High-achieving students: Their motivational goals, self-regulation and achievement and relationships to their teachers' goals and strategy-based instruction. *High Ability Studies, 14*(1), 23 – 39. https:/doi.org/10.1080/13598130304094

Erbas, A. K. & Bas, S. (2015). The contribution of personality traits, motivation, academic risk-taking and metacognition to the creative ability in mathematics. *Creativity Research Journal, 27*(4), 299 – 307. https://doi.org/10.1080/10400419.2015.1087235

Flores-Gonzalez, N. (2005). Popularity versus respect: School structure, peer groups and latino academic achievement. international *Journal of Qualitative Studies in Education (QSE), 18*(5), 625 – 642.

Forgasz, H. J. & Hill, J. C. (2013). Factors implicated in high mathematics achievement. *International Journal of Science and Mathematics Education, 11*(2), 481 – 499. https://doi.org/10.1007/s10763-012-9348-x

Gagné, F. (2016). From noncompetence to exceptional talent. Exploring the Range of Academic Achievement Within and Between Grade Levels. *Gifted Child Quarterly, 49*(2), 139 – 153. https://doi.org/10.1177/001698620504900204

Gagné, F. (1985). Giftedness and Talent: Reexamining a Reexamination of Definitions. *Gifted Child Quarterly, 29*(3), 103 – 112.

Galton, F. (1892). Hereditary genius. London: Macmillan and Co.

Geddes, K. A. (2011). Academic dishonesty among gifted and high-achieving students. *Gifted Child Today, 34*(2), 50 – 56. https://doi.org/10.1177/107621751103400214

Gough, D. (2007). Weight of evidence: a framework for the appraisal of the quality and relevance of evidence. *Research Papers in*

© 2022 The Author(s) Distributed as a Hogrefe OpenMind article under the license CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)

https://econtent.hogrefe.com/doi/pdf/10.1024/1010-0652/a000343 - Sunday, March 19, 2023 12:40:23 AM - Universitätsbibliothek Lüneburg IP Address:141.39.213.142

Education, 22(2), 213 – 228. https://doi.org/10.1080/02671520701296189

Griffin, K. A., Allen, W. R., Kimura-Walsh, E. & Yamamura, E. K. (2007). Those who left, those who stayed: Exploring the educational opportunities of high-achieving black and latina / o students at magnet and nonmagnet Los Angeles high schools (2001 – 2002). *Educational Studies: Journal of the American Educational Studies Association, 42*(3), 229 – 247.

Hany, E. A. (2012). Zum Verhältnis von Begabung und Leistung. In A. Hackl, C. Pauly, O. Steenbuck & G. Weigand (Hrsg.), *Werte schulischer Begabtenförderung. Begabung und Leistung (Karg-Hefte. Beiträge zur Begabtenförderung und Begabungsforschung)*, 35 – 40. Frankfurt, M.: Karg-Stiftung. https://doi.org/10.25656/01:9030

Harpalani, V. (2017). Counterstereotypic identity among high-achieving black students. *Penn GSE Perspectives on Urban Education, 14*(1), 1 – 9.

Heller, K. A. (Hrsg.). (2001). *Hochbegabung im Kindes- und Jugendalter* (2. Aufl.). Göttingen: Hogrefe.

Hochweber, J. (2010). *Was erfassen Mathematiknoten? Korrelate von Mathematik-Zeugniszensuren auf Schüler- und Schulklassenebene in Primar- und Sekundarstufe*. Waxmann.

Hoover, H. D., Dunbar, S. B. & Frisbie, D. A. (2001). *Iowa tests of basic skills*. Itasca, IL: Riverside Publishing.

Huang, H. & Zhu, H. (2017). High achievers from low socioeconomic backgrounds: The critical role of disciplinary climate and grit. *Mid-Western Educational Researcher, 29*(2), 93 – 116.

Ingenkamp, K. & Lissmann, U. (2008). *Lehrbuch der Pädagogischen Diagnostik*. Weinheim und Basel: Beltz Verlag.

Kaufman, S. B. (2009). Faith in intuition is associated with decreased latent inhibition in a sample of high-achieving adolescents. Psychology of Aesthetics, Creativity, and the Arts, 3(1), 28 – 34. https://doi.org/10.1037/a0014822

Köller, O. & Baumert, J. (2017). Hochleistende Schülerinnen und Schüler im mehr-und zweigliedrigen System. In M. Neumann, M. Becker, J. Baumert, K. Maaz & O. Köller (Hrsg.), *Zweigliedrigkeit im deutschen Schulsystem. Potenziale und Herausforderungen in Berlin* (1. Aufl., S. 227). Münster: Waxmann Verlag GmbH.

Kour, S. (2015). Scientific temper among academically high and low achieving adolescent girls. *Journal of Education and Practice, 6*(34), 96 – 101.

Krajewski, K., Dix, S. & Schneider, W. (2020). *Deutscher Mathematiktest für zweite Klassen* (2. Aufl.). Hogrefe.

Lauen, D. L. & Gaddis, S. M. (2016). Accountability pressure, academic standards, and educational triage. *Educational Evaluation and Policy Analysis, 38*(1), 127 – 147. https://doi.org/10.3102/0162373715598577

Lenhard, W., Lenhard, A. & Schneider, W. (2020). *Ein Leseverständnistest für Erst- bis Siebtklässler – Version II* (4. Aufl.). Göttingen: Hogrefe.

Lintorf, K. (2012a). Messtheoretische Güte von Schulnoten. In K. Lintorf (Hrsg.), *Wie vorhersagbar sind Grundschulnoten? Prädiktionskraft individueller und kontextspezifischer Merkmale* (S. 37 – 66). Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-94339-8_3

Lintorf, K. (Hrsg.). (2012b). *Wie vorhersagbar sind Grundschulnoten? Prädiktionskraft individueller und kontextspezifischer Merkmale*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Lubinski, D. (2016). From Terman to today. *Review of Educational Research, 86*(4), 900 – 944. https://doi.org/10.3102/0034654316675476

Lubinski, D. & Benbow, C. P. (2006). Study of Mathematically Precocious Youth After 35 Years: Uncovering Antecedents for the Development of Math-Science Expertise. *Perspectives on Psychological Science: a Journal of the Association for Psychological Science, 1*(4), 316 – 345. https://doi.org/10.1111/j.1745-6916.2006.00019.x

Lüftenegger, M., Kollmayer, M., Bergsmann, E., Jöstl, G., Spiel, C. & Schober, B. (2015). Mathematically gifted students and high achievement: The role of motivation and classroom structure. *High Ability Studies, 26*(2), 227 – 243. https://doi.org/10.1080/13598139.2015.1095075

Ma, X. (2005). A longitudinal assessment of early acceleration of students in mathematics on growth in mathematics achievement. *Developmental Review, 25*(1), 104 – 131. https://doi.org/10.1016/j.dr.2004.08.010

Maaz, K., Baeriswyl, F. & Trautwein, U. (2013). „Herkunft zensiert?" Leistungsdiagnostik und soziale Ungleichheiten in der Schule ["Origin graded?" Performance Diagnostics and Social Inequalities at School]. In D. Deißner (Hrsg.), *Chancen bilden. Wege zu einer gerechteren Bildung – ein internationaler Erfahrungsaustausch* (S. 185 – 188). Wiesbaden: Springer Fachmedien Wiesbaden.

Marsh, K. (2013). "Staying Black": The demonstration of racial identity and womanhood among a group of young high-achieving black women. *International Journal of Qualitative Studies in Education (QSE), 26*(10), 1213 – 1237. https://doi.org/10.1080/09518398.2012.731536

McBee, M. T. (2006). A descriptive analysis of referral sources for gifted identification screening by race and socioeconomic status. *Journal of Secondary Gifted Education, 17*(2), 103 – 111.

McGee, E. O. & Pearman, F. Alvin, II. (2015). Understanding black male mathematics high achievers from the inside out: Internal risk and protective factors in high school. *Urban Review: Issues and Ideas in Public Education, 47*(3), 513 – 540. https://doi.org/10.1007/s11256-014-0317-2

Mioduser, D. & Betzer, N. (2008). The contribution of project-based-learning to high-achievers' acquisition of technological knowledge and skills. *International Journal of Technology and Design Education, 18*(1), 59 – 77. https://doi.org/10.1007/s10798-006-9010-4

Möller, J. & Pohlmann, B. (2010). Achievement differences and self-concept differences. Stronger associations for above or below average students? *British Journal of Educational Psychology, 80*(3), 435 – 450. https://doi.org/10.1348/000709909X485234

Moser Opitz, E., Stöckli, M., Grob, U., Nührenbörger, M. & Reusser, L. (2019). *BASIS-MATH-G 3 + Gruppentest zur Basisdiagnostik Mathematik für das vierte Quartal der 3. Klasse und das erste Quartal der 4. Klasse*. Hogrefe.

Mourgues, C. V., Hein, S., Tan, M., Diffley III, R. & Grigorenko, E. L. (2016). The role of noncognitive factors in predicting academic trajectories of high school students in a selective private school. *European Journal of Psychological Assessment, 32*(1), 84 – 94. https://doi.org/10.1027/1015-5759/a000332

Neuendorf, C., Jansen, M. & Kuhl, P. (2020). Competence development of high achievers within the highest track in German secondary school: Evidence for Matthew effects or compensation? *Learning and Individual Differences, 77*, 101816. https://doi.org/10.1016/j.lindif.2019.101816

Neuendorf, C., Kuhl, P. & Jansen, M. (2017). Leistungsstarke Schülerinnen und Schüler in Deutschland. In P. Stanat, S. Schipolowski, C. Rjosk, S. Weirich & N. Haag (Hrsg.), *IQB-Bildungstrend 2016. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im zweiten Ländervergleich*. Münster: Waxmann.

Obergriesser, S. & Stoeger, H. (2016). The influence of emotions and learning preferences on learning strategy use before transition into high-achiever track secondary school. *High Ability Studies, 27*(1), 5 – 38. https://doi.org/10.1080/13598139.2015.1100980

© 2022 The Author(s) Distributed as a Hogrefe OpenMind article under the license CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)

https://econtent.hogrefe.com/doi/pdf/10.1024/1010-0652/a000343 - Sunday, March 19, 2023 12:40:23 AM - Universitätsbibliothek Lüneburg IP Address:141.39.213.142

OECD. (2016). *PISA 2015 Ergebnisse (Band 1). Exzellenz und Chancengerechtigkeit in der Bildung*. Bielefeld: PISA, W. Bertelsmann Verlag. https://doi.org/10.3278/6004573w

OECD. (2015). *The ABC of Gender Equality in Education. Aptitude, behaviour, confidence*. OECD Publishing. https://doi.org/10.1787/9789264229945-en

OECD. (2009). *Top of the class. High performers in science in PISA 2006*. OECD Publishing. https://doi.org/10.1787/9789264060777-en

Pant, H. A., Tiffin-Richards, S. P. & Köller, O. (2010). Standard-Setting für Kompetenztests im Large-Scale-Assessment. Projekt Standardsetting. *Zeitschrift für Pädagogik, 56*(Beiheft), 175 – 188.

Parsons, E. (2016). Does attending a low-achieving school affect high-performing student outcomes? *Teachers College Record, 118*(8), 1 – 36.

Peters, M. P. & Bain, S. K. (2011). Bullying and victimization rates among gifted and high-achieving students. *Journal for the Education of the Gifted, 34*(4), 624 – 643. https://doi.org/10.1177/016235321103400405

Preckel, F., Golle, J., Grabner, R., Jarvin, L., Kozbelt, A., Müllensiefen, D., et al. (2020). Talent development in achievement domains: A psychological framework for within- and cross-domain research. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 691 – 722. https://doi.org/10.1177/1745691619895030

Prenzel, M., Schütte, K. & Walter, O. (2007). Interesse an den Naturwissenschaften. In M. Prenzel, C. Artelt, J. Baumert, W. Blum & M. Hammann (Hrsg.), *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie* (S. 107 – 124). Münster: Waxmann.

Qin, D. B., Rak, E., Rana, M. & Donnellan, M. B. (2012). Parent-child relations and psychological adjustment among high-achieving Chinese and European American adolescents. *Journal of Adolescence, 35*(4), 863 – 873. https://doi.org/10.1016/j.adolescence.2011.12.004

Rambo-Hernandez, K. E. & McCoach, D. B. (2015). High-achieving and average students' reading growth. Contrasting school and summer trajectories. *The Journal of Educational Research, 108*(2), 112 – 129. https://doi.org/10.1080/00220671.2013.850398

Rathod, A. (2010). Self-regulated learning of high achievers. *Journal on Educational Psychology, 4*(2), 33 – 38.

Reilly, D., Neumann, D. L. & Andrews, G. (2015). Sex differences in mathematics and science achievement: A meta-analysis of National Assessment of Educational Progress assessments. *Journal of Educational Psychology, 107*(3), 645 – 662. https://doi.org/10.1037/edu0000012

Reiss, K., Weis, M. & Klieme, E. (2019). *PISA 2018. Grundbildung im internationalen Vergleich*.

Robinson, N. M., Lanzi, R. G., Weinberg, R. A., Ramey, S. L. & Ramey, C. T. (2002). Family factors associated with high academic competence in former Head Start children at third grade. *Gifted Child Quarterly, 46*(4), 278 – 290. https://doi.org/10.1177/001698620204600404

Roos, A.-L., Bieg, M., Goetz, T., Frenzel, A. C., Taxer, J. & Zeidner, M. (2015). Experiencing more mathematics anxiety than expected? Contrasting trait and state anxiety in high achieving students. *High Ability Studies, 26*(2), 245 – 258. https://doi.org/10.1080/13598139.2015.1095078

Rost, D. H. (Hrsg.). (2009). *Hochbegabte und hochleistende Jugendliche. Befunde aus dem Marburger Hochbegabtenprojekt* (Pädagogische Psychologie und Entwicklungspsychologie, Bd. 72, 2., erw. Aufl.). Münster: Waxmann.

Rost, D. H. & Buch, S. (2018). Hochbegabung. In D. H. Rost, J. R. Sparfeldt & S. Buch (Hrsg.), *Handwörterbuch pädagogische Psychologie* (Beltz Psychologie 2018, 5., überarbeitete und erweiterte Auflage, S. 226 – 242). Weinheim: Beltz.

Rothenbusch, S., Zettler, I., Voss, T., Lösch, T. & Trautwein, U. (2016). Exploring reference group effects on teachers' nominations of gifted students. *Journal of Educational Psychology, 108*(6), 883 – 897. https://doi.org/10.1037/edu0000085

Rüdiger, C., Jansen, M. & Rjosk, C. (2021). Empirische Arbeit: „Paul ist nicht so gut in Deutsch". Geschlechtsdifferenzielle Benotung im Fach Deutsch – eine Sekundäranalyse der Daten des IQB-Bildungstrends 2015. *Psychologie in Erziehung und Unterricht, 68*. https://doi.org/10.2378/peu2021.art08d

Rutkowski, D., Rutkowski, L. & Plucker, J. A. (2012). Trends in education excellence gaps: A 12-year international perspective via the multilevel model for change. *High Ability Studies, 23*(2), 143 – 166. https://doi.org/10.1080/13598139.2012.735414

Salmela, M. & Uusiautti, S. (2015). A positive psychological viewpoint for success at school – 10 characteristic strengths of the Finnish high-achieving students. *High Ability Studies, 26*(1), 117 – 137. https://doi.org/10.1080/13598139.2015.1019607

Schmidt, F. L. & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*(2), 262 – 274. https://doi.org/10.1037/0033-2909.124.2.262

Schurtz, I. M., Pfost, M. & Artelt, C. (2014). Variieren die Selbstkonzeptdifferenzen in Abhängigkeit vom Leistungsniveau? Differenzielle Zusammenhänge in Deutsch, Englisch und Mathematik. *Zeitschrift für Pädagogische Psychologie, 28*(1 – 2), 31 – 42. https://doi.org/10.1024/1010-0652/a000122

Schwippert, K., Kasper, D., Köller, O., McElvany, N., Selter, C., Steffensky, M. et al. (Hrsg.). (2020). *TIMSS 2019. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (1. Auflage). Münster: Waxmann.

Simons, D. J., Shoda, Y. & Lindsay, D. S. (2017). Constraints on Generality (COG): A Proposed addition to all empirical papers. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science, 12*(6), 1123 – 1128. https://doi.org/10.1177/1745691617708630

Simonsohn, U., Simmons, J. P. & Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behaviour, 4*(11), 1208 – 1214. https://doi.org/10.1038/s41562-020-0912-z

Sontag, C. & Stoeger, H. (2015). Can highly intelligent and high-achieving students benefit from training in self-regulated learning in a regular classroom context? *Learning and Individual Differences, 41*, 43 – 53. https://doi.org/10.1016/j.lindif.2015.07.008

Sparfeldt, J. R., Buch, S. R. & Rost, D. H. (2010). Klassenprimus bei durchschnittlicher Intelligenz. *Zeitschrift für Pädagogische Psychologie, 24*(2), 147 – 155. https://doi.org/10.1024/1010-0652/a000012

Stanat, P., Schipolowski, S., Mahler, N., Weirich, S. & Henschel, S. (Hrsg.). (2019). *Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I im zweiten Ländervergleich*. Münster: Waxmann.

Steegen, S., Tuerlinckx, F., Gelman, A. & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science, 11*(5), 702 – 712. https://doi.org/10.1177/1745691616658637

Steinmayr, R. & Spinath, B. (2017). Why time constraints increase the gender gap in measured numerical intelligence in academically high achieving samples. European Journal of Psychological Assessment. https://doi.org/10.1027/1015-5759/a000400

Stoeger, H., Hopp, M. & Ziegler, A. (2017). Online mentoring as an extracurricular measure to encourage talented girls in STEM

© 2022 The Author(s) Distributed as a Hogrefe OpenMind article under the license CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)

(science, technology, engineering, and mathematics): An empirical study of one-on-one versus group mentoring. *Gifted Child Quarterly, 61*(3), 239 – 249. https://doi.org/10.1177/0016986217702215

Subotnik, R. F., Olszewski-Kubilius, P. & Worrell, F. C. (2011). Rethinking giftedness and gifted education: A proposed direction forward based on psychological science. *Psychological Science in the Public Interest: A Journal of the American Psychological Society, 12*(1), 3 – 54. https://doi.org/10.1177/1529100611418056

Terman, L. M. (1954). The discovery and encouragement of exceptional talent. *American Psychologist, 9*, 221 – 230. https://doi.org/10.1037/h0060516

Terman, L. M. (1926). *Genetic studies of genius. Mental and physical traits of a thousand gifted children*. Stanford University Press.

Trapmann, S., Hell, B., Weigand, S. & Schuler, H. (2007). Die Validität von Schulnoten zur Vorhersage des Studienerfolgs – eine Metaanalyse. *Zeitschrift für Pädagogische Psychologie 21*(1), 11 – 27. https://doi.org/10.1024/1010-0652.21.1.11

Trautwein, U., Köller, O. & Kämmerer, E. (2002). Effekte innerer und äußerer Leistungsdifferenzierung auf selbstbezogene Fähigkeitskognitionen, die wahrgenommene Unterrichtspartizipation und die wahrgenommene soziale Akzeptanz. *Psychologie in Erziehung und Unterricht, 49*(4), 273 – 286.

Vock, M., Köller, O. & Nagy, G. (2013). Vocational interests of intellectually gifted and highly achieving young adults. *British Journal of Educational Psychology, 83*(2), 305 – 328. https://doi.org/10.1111/j.2044-8279.2011.02063.x

Walker, C. L. & Shore, B. M. (2015). Myth busting: Do high-performance students prefer working alone? *Gifted and Talented International, 30*(1 – 2), 85 – 105. https://doi.org/10.1080/15332276.2015.1137461

Wang, M.-T., Eccles, J. S. & Kenny, S. (2013). Not lack of ability but more choice: individual and gender differences in choice of careers in science, technology, engineering, and mathematics. *Psychological Science, 24*(5), 770 – 775. https://doi.org/10.1177/0956797612458937

Weckbacher, L. M. & Okamoto, Y. (2012). Spatial experiences of high academic achievers: insights from a developmental perspective. *Journal for the Education of the Gifted, 35*(1), 48 – 65. https://doi.org/10.1177/0162353211432038

Woodcock, R. W., McGrew, K. S. & Mather, N. (2001). *Woodcock-Johnson III tests of achievement*. Itasca, IL: Riverside Publishing.

Yong, B. C. S. (2012). Comparison between the thinking styles of students in a science school and a mainstream school. *Journal of Science and Mathematics Education in Southeast Asia, 35*(1), 60 – 83.

Zhou, Y., Fan, X., Wei, X. & Tai, R. H. (2017). Gender gap among high achievers in math and implications for STEM pipeline. *Asia-Pacific Education Researcher, 26*(5), 259 – 269. https://doi.org/10.1007/s40299-017-0346-1

Ziegler, A. & Raul, T. (2000). Myth and Reality: A review of empirical studies on giftedness. *High Ability Studies, 11*(2), 113 – 136. https://doi.org/10.1080/13598130020001188

Zimmer, K., Brunner, M., Lüdtke, O., Prenzel, M. & Baumert, J. (2007). Die PISA-Spitzengruppe in Deutschland: Eine Charakterisierung hochkompetenter Jugendlicher. In K. A. Heller & A. Ziegler (Hrsg.), *Begabt sein in Deutschland* (Talentförderung, Expertiseentwicklung, Leistungsexzellenz, Bd. 1, S. 193 – 208). Berlin: Lit.

Historie
Manuskript eingereicht: 02.07.2021
Manuskript nach Revision angenommen: 27.01.2022
Onlineveröffentlichung: 10.02.2022

Danksagung
Hiermit danken wir Nikolas Weigt für seine Unterstützung bei der Kodierung der Studien.

Förderung
Open-Access-Veröffentlichung ermöglicht durch die Eberhard Karls Universität Tübingen.

ORCID
Claudia Neuendorf
https://orcid.org/0000-0002-3024-0000
Malte Jansen
https://orcid.org/0000-0001-7081-6505
Poldi Kuhl
https://orcid.org/0000-0001-8021-8840
Miriam Vock
https://orcid.org/0000-0003-0101-1003

Claudia Neuendorf
Eberhard Karls Universität Tübingen
Hector-Institut für Empirische Bildungsforschung
Geschwister-Scholl-Platz
72074 Tübingen
Deutschland

claudia.neuendorf@uni-tuebingen.de

https://econtent.hogrefe.com/doi/pdf/10.1024/1010-0652/a000343 - Sunday, March 19, 2023 12:40:23 AM - Universitätsbibliothek Lüneburg IP Address:141.39.213.142

© 2022 The Author(s) Distributed as a Hogrefe OpenMind article under the license CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)

Anhang

Tabelle A1. Suchstrategie bei der Literaturrecherche in Fachdatenbanken

Datenbank	Hauptsuchbegriffe	Filter	*N*
PsychARTICLES	„high-achiev*“		55
	“high-achiev*” OR “high achiev”* AND students		4
Psyndex Literature & AV Media	leistungsstark*	TYPE: Academic Journals	52
ERIC	high achievers AND students	Filter: • wissenschaftliche Zeitschriften • high achievement (Thema) • NOT(Early Childhood Education, Higher Education, Kindergarten, Two Year Colleges) (Bildungsetappen)	146
	high performing students	Filter: • wissenschaftliche Zeitschriften • high achievement (Thema) • NOT(Early Childhood Education, Higher Education, Kindergarten, Two Year Colleges) (Bildungsetappen)	28

Tabelle A2. Kurzüberblick über eingeschlossene Studien

Autor, Jahr	N	Alters-gruppe	Forschungs-design	Land	Indikatoren	Fachbezug	Bezugsnorm
Adelodun et al., 2015	50	O	quantitativ	Nigeria	Noten	Einzelfach	-
Allen & Griffin, 2006	17	M – O	qualitativ	USA	Noten, Kontext	Gesamtwert	kriterial
Assouline et al., 2017	1146	M	experimentell	USA	Test	Gesamtwert	sozial (P)
Bütüner & Filiz, 2017	233	G	qualitativ	Türkei	Noten	Einzelfach	kriterial
Barron, 2000	96	M	experimentell	USA	Schulform	-	kriterial
Bergold et al., 2017	74868	G	quantitativ	Multinational (EU)	LSA	Kombination	kriterial
Bergold et al., 2020	3928	G	quantitativ	Deutschland	LSA	Kombination	sozial (P)
Berkowitz & Cicchelli, 2004	63	M	quantitativ	USA	Noten	Einzelfach	kriterial
Carroll & Bailey, 2016	967	G	quantitativ	USA	LSA	Kombination	kriterial
Carter Andrews, 2012	2181	M – O	qualitativ	USA	Noten, Einschätzung, Kontext, Andere	Gesamtwert	sozial (S)
Carter, 2008	9	O	qualitativ	USA	Noten, Kontext, Andere	Gesamtwert	kriterial
Castejón & Zancajo, 2015	515655	M	quantitativ	Multinational	LSA	Einzelfach	sozial (P)
Cheung, 2017	22636	M	quantitativ	Multinational (Asien)	LSA	Einzelfach	sozial (P)
Clausen et al., 2013	53	M	qualitativ	Deutschland	Schulform	-	kriterial
Crawford et al., 2017	480653	G	quantitativ	Großbritannien	LSA	Einzelfach	kriterial
Ee et al., 2003	566	G	quantitativ	Singapur	Schulform	-	sozial (P)
Erbas & Bas, 2015	217	M	quantitativ	Türkei	Schulform	Kombination	kriterial
Flores-Gonzalez, 2005		O	qualitativ	USA	Noten, Kontext	Gesamtwert	kriterial
Forgasz & Hill, 2013	4811	O	qualitativ	Australien	Noten	Einzelfach	sozial (P)
Gagné, 2016 (2)		G – M	quantitativ	USA	Test	Gesamtwert	sozial (P)
Geddes, 2011		M – O	quantitativ	USA	Kursstufe	Kombination	kriterial
Griffin et al., 2007	34	O	qualitativ	USA	Noten, Kontext	Gesamtwert	kriterial
Harpalani, 2017	779	M	quantitativ	USA	Noten	Gesamtwert	kriterial

https://econtent.hogrefe.com/doi/pdf/10.1024/1010-0652/a000343 - Sunday, March 19, 2023 12:40:23 AM - Universitätsbibliothek Lüneburg IP Address:141.39.213.142

© 2022 The Author(s) Distributed as a Hogrefe OpenMind article under the license CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)

Tabelle A2. Kurzüberblick über eingeschlossene Studien (Fortsetzung)

Autor, Jahr	N	Alters-gruppe	Forschungs-design	Land	Indikatoren	Fachbezug	Bezugsnorm
Huang & Zhu, 2017 (2)	1220	M	quantitativ	USA	LSA	Einzelfach	sozial (P)
Kaufman, 2009	162	O	experimentell	Großbritannien	Schulform	Gesamtwert	kriterial
Kour, 2015	120	O	quantitativ	Indien	Noten	-	kriterial
Lauen & Gaddis, 2016 (2)	500000	G – M	quantitativ	USA	LSA	Einzelfach	sozial (P)
Lüftenegger et al., 2015	210	M – O	quantitativ	Österreich	Noten	Einzelfach	kriterial
Ma, 2005	3116	M	quantitativ	USA	Test	Einzelfach	sozial (S)
Marsh, 2013	9	O	qualitativ	USA	Schulform	Kombination	kriterial
McGee et al., 2015	13	M – O	qualitativ	USA	Kursstufe	Einzelfach	sozial (K)
Mioduser & Betzner, 2008	120	O	experimentell	Israel	Noten	-	-
Mourgues et al., 2016	8586	M	quantitativ	USA	Schulform, Noten	Gesamtwert	sozial (S)
Neuendorf et al., 2020 (2)	1010	M	quantitativ	Deutschland	Test	Einzelfach	sozial (S)
Obergriesser & Stoeger, 2016	200	G	quantitativ	Deutschland	Noten	Mittelwert	kriterial
Parsons, 2016		G	quantitativ	USA	LSA	Einzelfach	sozial (P)
Peters & Bain, 2011	90	M – O	quantitativ	USA	Kursstufe	-	sozial (K)
Qin et al., 2012	487	M	quantitativ	USA	Schulform	-	kriterial
Rambo-Hernandez & McCoach, 2015 (1)	21021	G	quantitativ	USA	Test	Einzelfach	sozial (K)
Rambo-Hernandez & McCoach, 2015 (2)	70521	G	quantitativ	USA	Test	Einzelfach	sozial (P)
Rathod, 2010	480	O	quantitativ	Indien	Noten	-	kriterial
Reilly et al., 2015 (2)		G – O	quantitativ	USA	LSA	Einzelfach	kriterial
Robinson et al., 2002	5400	G	mixed methods	USA	Test	Kombination	sozial (S)
Roos et al., 2015	237	M	quantitativ	Deutschland	Noten	Einzelfach	kriterial
Rutkowski et al., 2012 (2)	272000	M	quantitativ	Multinational	LSA	Einzelfach	kriterial
Salmela & Uusiautti, 2015	14	O	qualitativ	Finnland	Noten	Gesamtwert	kriterial
Sontag & Stoeger, 2015	123	G	experimentell	Deutschland	Noten	Mittelwert	sozial (S)
Sparfeldt et al. 2010	256	M	quantitativ	Deutschland	Noten, Lehrkräfte-einschätzungen	Mittelwert	kriterial
Steinmayr & Spinath, 2017	666	O	quantitativ	Deutschland	Schulform	-	kriterial
Stoeger et al., 2017	347	M – O	experimentell	Deutschland	Schulform	-	kriterial
Vock et al., 2013	4694	O	quantitativ	Deutschland	Noten	Gesamtwert	sozial (S)
Walker & Shore, 2015	69	G	quantitativ	Kanada	Schulform	-	kriterial
Wang et al., 2013	1490	O	quantitativ	USA	Test	Kombination	sozial (S)
Weckbacher & Okamoto, 2012	43	M	quantitativ	USA	Noten, Test, Kontext	Kombination	sozial (P)
Yong, 2012	378	M	quantitativ	Brunei	Schulform	Einzelfach	kriterial
Zhou et al., 2017 (4)	130400	M	quantitativ	Multinational	LSA	Einzelfach	sozial (P)

Anmerkungen: Eine vollständige Übersicht über die eingeschlossenen Studien sowie die Kodierung aller im Artikel verwendeter Merkmale wird unter https://osf.io/jzkv6/ bereitgestellt.
Altersgruppe: O = Oberstufe / Sekundarstufe II, M = Mittelstufe / Sekundarstufe I, G = Grundschule. LSA = Large Scale Assessment. P = Populationsebene, S = Stichprobenebene, K = Klassenebene.

https://econtent.hogrefe.com/doi/pdf/10.1024/1010-0652/a000343 - Sunday, March 19, 2023 12:40:23 AM - Universitätsbibliothek Lüneburg IP Address:141.39.213.142

© 2022 The Author(s) Distributed as a Hogrefe OpenMind article under the license CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)

Tabelle A3. In den Studien verwendete Notensysteme und Noten-Cut-offs

Land	Notensystem	Cut-off-Werte
USA	GPA	2.8[a]; 3[b]; 4[c]
	0 – 100 %	95 %[c]
	Letter scale F-A	B[d]; C[e]
Australien	0 – 50 Punkte	46[f]
Indien	0 – 100 %	60 %[g]; 70 %[h]
Türkei	1 – 5	4[i]
Österreich	5 – 1	1[j]
Deutschland	6 – 1	1.4[k]; 1.75[l]; 2.33[m]
Finnland	Letter scale I-L	L[n]

Anmerkung: Dargestellt sind die in den Studien genutzten Notenwerte und ihre Cut-offs. Die Werte basieren auf 17 Studien. [a] Carter Andrews, 2012; Carter, 2008. [b] Allen & Griffin, 2006; Griffin et al., 2007. [c] Weckbacher & Okamoto, 2012; Berkowitz & Chicchelli, 2004. [d] Harpalani, 2017. [e] Flores-Gonzales, 2005. [f] Forgasz & Hill, 2013. [g] Rahtod, 2010. [h] Kour, 2015. [i] Bütüner & Filiz, 2017. [j] Lüftenegger, Kollmayer, Bergsmann, Jöstl, Spiel & Schober, 2015. [k] Sparfeldt, Buch & Rost, 2010. [l] Roos, Bieg, Goetz, Frenzel, Taxer & Zeidner, 2015. [m] Obergriesser & Stoeger, 2016. [n] Salmela & Uusiautti, 2015.

https://econtent.hogrefe.com/doi/pdf/10.1024/1010-0652/a000343 - Sunday, March 19, 2023 12:40:23 AM - Universitätsbibliothek Lüneburg IP Address:141.39.213.142

© 2022 The Author(s) Distributed as a Hogrefe OpenMind article under the license CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)