

Schüssler-Hanenberg, Sarah

**Leistungsbewertung im naturwissenschaftlich-technischen Sachunterricht der Grundschule. Eine Untersuchung der Leistungsbewertung technischer Artefakte deutschstämmiger und immigrierter Schülerinnen und Schüler**

*technik-education (tedu). Fachzeitschrift für Unterrichtspraxis und Unterrichtsforschung im allgemeinbildenden Technikunterricht 3 (2023) 2, S. 21-28*



Quellenangabe/ Reference:

Schüssler-Hanenberg, Sarah: Leistungsbewertung im naturwissenschaftlich-technischen Sachunterricht der Grundschule. Eine Untersuchung der Leistungsbewertung technischer Artefakte deutschstämmiger und immigrierter Schülerinnen und Schüler - In: *technik-education (tedu). Fachzeitschrift für Unterrichtspraxis und Unterrichtsforschung im allgemeinbildenden Technikunterricht 3 (2023) 2, S. 21-28* - URN: urn:nbn:de:0111-pedocs-286477 - DOI: 10.25656/01:28647

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-286477>

<https://doi.org/10.25656/01:28647>

in Kooperation mit / in cooperation with:



<https://tec-edu.net/tedu>

**Nutzungsbedingungen**

Dieses Dokument steht unter folgender Creative Commons-Lizenz: <http://creativecommons.org/licenses/by-nc/4.0/deed.de> - Sie dürfen das Werk bzw. den Inhalt vervielfältigen, verbreiten und öffentlich zugänglich machen sowie Abwandlungen und Bearbeitungen des Werkes bzw. Inhaltes anfertigen, solange Sie den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen und das Werk bzw. den Inhalt nicht für kommerzielle Zwecke verwenden.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

**Terms of use**

This document is published under following Creative Commons-Licence: <http://creativecommons.org/licenses/by-nc/4.0/deed.en> - You may copy, distribute and render this document accessible, make adaptations of this work or its contents accessible to the public as long as you attribute the work in the manner specified by the author or licensor. You are not allowed to make commercial use of the work, provided that the work or its contents are not used for commercial purposes.

By using this particular document, you accept the above-stated conditions of use.



**Kontakt / Contact:**

peDOCS  
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation  
Informationszentrum (IZ) Bildung  
E-Mail: [pedocs@dipf.de](mailto:pedocs@dipf.de)  
Internet: [www.pedocs.de](http://www.pedocs.de)

# technik – education

3. Jahrgang

Fachzeitschrift für Unterrichtspraxis und Unterrichtsforschung  
im allgemeinbildenden Technikunterricht

2|2023



[www.tec-edu.net](http://www.tec-edu.net)

**tedu**

# Fachzeitschrift für Unterrichtspraxis und Unterrichtsforschung im allgemeinbildenden Technikunterricht

<https://tec-edu.net/tedu>

## HERAUSGEBER

Dr. Hannes Helmut Nepper  
Dr. Armin Ruch, OStR  
Dr. Dr. Dierk Suhr

## Mail

herausgeber@tec-edu.net

## Anschrift

Pädagogische Hochschule Schw. Gmünd  
Institut für Bildung, Beruf und Technik  
Abteilung Technik  
Oberbettringer Straße 200  
73525 Schwäbisch Gmünd  
[www.tec-edu.net](http://www.tec-edu.net)

## AUTOR\*INNEN IN DIESEM HEFT

Daniel Autenrith  
Fabian Csoch  
Sebastian Rudolf Göser  
Niclas Günther  
Yannik Haußmann  
Nils Heyden  
Stefanie Nickel  
Armin Ruch  
Sarah Schüssler-Hanenberg

Titelbild: Hannes Helmut Nepper  
(Nutzung von Adobe Firefly)

**ISSN: 2748-2022**

## Inhalt

**Grußwort der Herausgeber ..... 2**

*Unterrichtsforschung*

N. Heyden

**Präkonzepte zum Kohlekraftwerk ..... 3**

*Unterrichtsforschung*

D. Authenrieth & S. Nickel

**Das KI-Meta-Modell ..... 14**

*Unterrichtsforschung*

S. Schüssler-Hanenberg

**Leistungsbewertung ..... 21**

*Unterrichtspraxis*

Y. Haußmann

**Fertigung von Ringen im Technikunterricht ..... 29**

*Unterrichtspraxis*

N. Günther

**Schachbrett für Menschen mit Sehbehinderung ..... 35**

*Unterrichtspraxis*

S. R. Göser

**Ein Bürolocher aus dem 3D-Drucker ..... 46**

*Unterrichtspraxis*

F. Csoch

**Umwelt- und Nachhaltigkeitsaspekte im Technikunterricht ..... 56**

*Ankündigungen*

A. Ruch

**Neue Fachliteratur ..... 63**

Namentlich gekennzeichnete Beiträge geben nicht unbedingt die Meinung der Herausgeber wieder.

Insbesondere bei unterrichtspraktischen Artikeln wird darauf hingewiesen, dass es unterschiedliche Sicherheitsbestimmungen gibt und jede Lehrkraft bei der Umsetzung selbst dafür verantwortlich ist, die Gefährdung zu beurteilen und die Vorschläge für die eigene Praxis entsprechend der jeweilige Vorschriftenlage anzupassen.

# Leistungsbewertung im naturwissenschaftlich-technischen Sachunterricht der Grundschule

## Eine Untersuchung der Leistungsbewertung technischer Artefakte deutschstämmiger und immigrierter Schülerinnen und Schüler

*Sarah Schüssler-Hanenberg*

---

### SCHLAGWORTE

*Diskriminierung  
Leistungsbewertung  
Sachunterricht  
Migrationshintergrund*

### ABSTRACT

Stereotype führen zu verzerrter Wahrnehmung von Menschengruppen. In der Schule kann es so zu Verzerrungen bei der Beurteilung von Schülerleistungen kommen. In dieser Studie wurde der Einfluss eines Migrationshintergrunds auf die Vergabe von Schulnoten durch angehende Lehrkräfte (N = 44) untersucht. Die Probanden bewerteten dieselbe Leistung eines Lernenden mit Migrationshintergrund im Durchschnitt etwas schlechter als die eines Lernenden ohne Migrationshintergrund. Der Unterschied war bei den schlechteren Arbeiten ausgeprägter. Interessanterweise ergaben sich Unterschiede in der Bewertung einzelner Schülernamen, wobei der Name ‚Murat‘ am stärksten benachteiligt wurde. Außerdem wurden Defizite in der Lehrerausbildung aufgezeigt, insbesondere in Bezug auf die Bewertung von technischen Artefakten und die einheitliche Anwendung von Bewertungskriterien. Die Ergebnisse stützen die Annahme, dass angehende Lehrkräfte bei der Benotung von Schülern mit Migrationshintergrund voreingenommen sind und dass sich dies negativ auf die Beurteilung von Schülern mit Migrationshintergrund auswirkt.

### Einleitung

Unterschiede in der Benotung von Lernenden mit und ohne Migrationshintergrund werden auf verschiedene Ursachen zurückgeführt. Es wurde nachgewiesen, dass immigrierte Lernende Schwierigkeiten in der Leistungsfähigkeit aufgrund von bspw. Sprachschwierigkeiten oder ungünstigen häuslichen Förderungsbedingungen aufweisen und dementsprechend tatsächlich schlechtere Leistungen erbringen können als Lernende ohne Migrationshintergrund. Zusätzlich haben sie Nachteile, die durch das Bildungssystem oder die Lehrperson verursacht werden. Eine gegenüber Lernenden mit Migrationshintergrund voreingenommene Lehrkraft neigt eher dazu diese schlechter zu bewerten und weniger von ihnen zu erwarten als von Lernenden ohne Migrationshintergrund. In der vorliegenden Querschnittstudie wurde sich für den Vergleich zwischen deutschen und türkischen Lernenden entschieden, da türkische Migranten die größte Migrantengruppe in Deutschland und Baden-Württemberg bilden. Außerdem haben frühere Studien gezeigt, dass diese Gruppe mit den meisten Stereotypen verbunden wird (u.a. Kahraman & Knoblich 2000). Ebenso stellen sie in Bezug auf Benachteiligung in der Schule die wichtigste Gruppe dar.

Da Schulnoten ein wichtiger Faktor für den Bildungserfolg darstellen, die Bewertung von Arbeiten durch Lehrkräfte aber nicht fehlerfrei ist, wurde im Rahmen dieser Studie untersucht, ob es Bewertungsunterschiede zwischen Arbeiten von deutschen und türkischen Lernenden gibt. Es sollte

untersucht werden, ob bei der Bewertung von Lernenden mit Migrationshintergrund bewusste oder unbewusste Stereotype eine Rolle spielen.

Um dies zu klären, wurden Studierende des Grundschullehramts, die den Schwerpunkt Sachunterricht haben, interviewt: ihnen wurden vier Werkstücke (Untersetzer) in zwei verschiedenen Qualitäten gezeigt, welche von ihnen benotet werden sollten. Die Qualität der Lernenden wurde dabei zwischendurch variiert: die Hälfte der Probanden benotete eine gute Leistung eines Schülers, der entweder deutscher oder türkischer Herkunft zu sein schien, die andere Hälfte der Probanden bewertete eine durchschnittliche Leistung, die von einem Schüler, der entweder deutscher oder türkischer Herkunft schien, erbracht worden war. Dies geschah, um die durchschnittliche Leistung eines Artefaktes als Kontrollvariable festzustellen und um daraus Schlüsse auf die Leistungsbewertung einzelner Schüler ziehen zu können. Die Werkstücke wurden im Vorfeld zwei deutschen und zwei türkischen Schülern zugeordnet und ihre Namen mit Alter und Klassenstufe auf den Werkstücken angebracht. Im Anschluss an die Benotung der Werkstücke sollten die Probanden ihre gegebenen Noten anhand der eventuell zur Hilfe genommen Kriterien begründen. Daraus können Schlüsse auf ihr Vorgehen bei der Notenfindung und auf ihre Haltung den Schülern gegenüber gezogen werden. Da nicht direkt nach Vorurteilen oder bekannten Stereotypen gefragt wurde, konnte nur aus dem Gesagten und dem Kontext auf

eventuelle Berücksichtigung dieser bei der Benotung geschlossen werden. Hierbei spielten besonders die Aussagen aus Frage 3 eine Rolle, bei denen die Probanden hypothetische Beispiele aus dem vermeintlichen Fertigungsprozess nannten (z. B. „Wenn jetzt natürlich klar ist, dass, ähm, keine Ahnung, ähm, Murat letztlich überall Hilfe gebraucht hat...“; vgl. Interview 1). Die Untersuchung von Kahraman und Knoblich (2000) und die Beobachtungen von Nickerson (1998) lassen darauf schließen, dass Lehrkräfte ein auf Stereotypen basiertes Bild von türkischstämmigen Lernenden haben und dieses in ihre Bewertung von Arbeiten dieser mit einfließen lassen könnten.

Die Probanden gaben den einzelnen Artefakten Schulnoten (1,00 bis 6,00) und beschrieben anhand der Artefakte die guten bzw. weniger guten Leistungen, die sie ausmachen konnten. Dies sind beides subjektive Herangehensweisen, da die Probanden keinen Kriterienkatalog zur Berücksichtigung hatten und deshalb ihre eigenen Kriterien festlegen konnten. Dadurch sollte herausgefunden werden, ob die Probanden ihre herangezogenen Kriterien für alle Schüler gleich hielten oder ob sich die Kriterien oder ihre Gewichtung für Schüler mit bzw. ohne Migrationshintergrund unterschieden.

## BISHERIGE FORSCHUNG

Bisherige Studien haben gezeigt, dass die Lehrkraft eine entscheidende Rolle in der Gestaltung der Bildungsverläufe von Lernenden hat. Neben ihren Entscheidungen über Noten und Förderung und ihren Empfehlungen – die nicht bindend sind – in Bezug auf den Besuch einer weiterführenden Schule, hat die Interaktion mit den Lernenden im Unterricht Auswirkungen auf die Leistung dieser. Jussim und Harber (2005) konnten herausfinden, dass Lehrkräfte unterschwerlich die Lernenden beeinflussen und so selbsterfüllende Prognosen auslösen können. Besonders sind Lernende aus stigmatisierten sozialen Gruppen davon betroffen: die (herabgesetzten) Erwartungen der Lehrkraft an Leistung und Verhalten der Lernenden führen dazu, dass diese die Rolle einnehmen, die die Lehrkraft ihnen zuweist. Lehrkräfte tragen also direkt und indirekt zu Unterschieden zwischen einzelnen Lernenden bei. Außerdem bestätigten Walton und Spencer (2009) in ihrer Studie über das Abschneiden von Lernenden in Tests, dass diejenigen, die durch die Lehrkraft stigmatisiert werden, in Tests besser abschneiden, wenn die Lehrperson nicht die Person ist, die den Test im Nachhinein bewertet.

Lehrkräfte berücksichtigen nicht alle Informationen über einen Lernenden bei der Bewertung, werden aber durchaus durch bekannte Stereotype und Vorurteile, also weitgehend gemeinsame Annahmen in einer Gesellschaft über Merkmale von Mitgliedern einer Gruppe, beeinflusst. Stereotype können bedingen, dass die Lehrkräfte geringere Erwartungen an die Arbeiten der Lernenden mit Migrationshintergrund haben. Dies variiert von Lehrkraft zu Lehrkraft, je nach ihren persönlichen Einstellungen. Kahraman und Knoblich (2000) haben in einer Untersuchung zu kulturellen Stereotypen in Deutschland nachgewiesen, dass Türken im Vergleich zu

Deutschen ohne Migrationshintergrund weniger mit Leistung und Erfolg in Verbindung gebracht werden. Außerdem konnten sie feststellen, dass die Probanden gleich viele Stereotype zu Deutschen und Türken nennen, die Stereotype über Türken aber viel seltener positiv ausfallen. 2015 veröffentlichten Glock et al. ihre Studie, bei der der Einfluss von Stereotypen auf Lehrkräfte untersucht wurde. Wurden den Lehrkräften Informationen über einen immigrierten Lernenden genannt, die gängige Stereotype zu seiner Nationalität, z. B. zum sozioökonomischen Status, bedienten, so ließen sie sich davon bei ihrer Beurteilung dieses Lernenden mehr beeinflussen als Lehrkräfte, denen keine Informationen gegeben wurden. Außerdem konnte aufgezeigt werden, dass dieselben Leistungen in Abhängigkeit verschiedener Merkmale, wie z.B. dem sozioökonomischen Status eines Lernenden, unterschiedliche Bewertungen erhalten können: so wurde in einer Studie von Darley und Gross (1983) nachgewiesen, dass wenn Befragte Informationen über den sozialen Status eines Lernenden hatten, sie die Arbeiten von Kindern mit höheren sozialen Status besser bewerteten, als Befragte, die Arbeiten von Kindern mit vermeintlich geringerem sozialen Status bewerteten (Nickerson 1998, S. 182).

Nickerson (1998) spricht davon, dass Personen sich von ihren Erwartungen beeinflussen lassen: Menschen tendieren dazu, ihre eigenen Erwartungen selbst zu bestätigen. So können Lehrpersonen unbewusst eine Leistung eines stigmatisierten Lernenden schlechter beurteilen und Lernende dahingehend beeinflussen, dass sie dem vorurteilsbehafteten Bild der Lehrkraft entsprechen.

In den Studien von Bonefeld et al. 2017 bzw. 2018 wurde zum einen im Fach Deutsch die Benotung von Diktaten vermeintlich türkischer und deutscher Drittklässler untersucht. Sie konnten nachweisen, dass Lehrkräfte in Ausbildung die Leistungen immigrierter Schüler\*innen signifikant schlechter benoteten als dieselbe Leistung eines deutschen Lernenden (Bonefeld & Dickhäuser 2018, S. 7f). Im Fach Mathematik an deutschen Gymnasien beobachteten Bonefeld et al. (2017), dass Lernende mit Migrationshintergrund signifikant schlechtere Klassenarbeits- und Zeugnisnoten erhielten, auch wenn standardisierte Tests als Kontrolle durchgeführt wurden. Sie untersuchten dafür die Leistungsbeurteilungen von Lernenden von Beginn der fünften bis zum Ende der sechsten Klasse, wodurch aufgezeigt werden konnte, dass dieser Effekt im Laufe der Zeit stabil blieb (Bonefeld et al. 2017, S. 16ff).

## METHODE

Im Folgenden werden die Methoden vorgestellt.

### Entwicklung und Vorprüfung des Materials

Um zu testen, ob die Leistungsbewertung der Probanden dem tatsächlichen Leistungsniveau der angeblichen Schüler entspricht, wurden vier Untersetzer aus Holz von der Versuchsleitung gebaut, die zwei unterschiedliche Leistungsniveaus aufweisen. Es sollte untersucht werden, ob es durch die Probanden zu Unterschieden in der Bewertung von Artefakten zwischen Schülern mit und ohne Migra-

tionshintergrund kommt. Deshalb wurden für alle Schüler dieselbe Klassenstufe und dasselbe Alter gewählt. Um geschlechtsspezifische Unterschiede auszuschließen, wurden ausschließlich männliche Namen verwendet.

#### Artefakte zur Operationalisierung des Leistungsniveaus

Um die Hypothesen zu testen, wurden vier Topfuntersetzer (Artefakte A bis D) aus Holz gefertigt. Sie bestanden aus vier Holzlatten, sechs Holzzyllindern und einem Aluminiumdraht. Für die Studie wurden zwei Artefakte gewählt, die ein objektiv gutes Leistungsniveau aufweisen und zwei Artefakte, die ein objektiv durchschnittliches Leistungsniveau darstellen: zwei Artefakte (A und B) wiesen nur geringe Mängel auf und die anderen zwei Artefakte (C und D) hatten jeweils neben kleineren Mängeln einen groben Fehler. Sie wurden zunächst von Studierenden des Sekundarlehramts mit Schwerpunkt Technik (N = 4) ohne Informationen über die vermeintlichen Schüler, außer dass sie in der 3. Klasse waren, in Leistungsgruppen eingeteilt, um zu validieren, dass die Artefakte zwei verschiedene Leistungsniveaus aufweisen. Das Ergebnis dieser Validierung war eindeutig: es gab zwei Leistungsniveaus, die durch die groben Fehler unterschieden wurden. Die Befragten waren sich einig, dass die Arbeiten zu ordentlich waren, um von Schülern der 3. Klasse hergestellt worden zu sein. Daraufhin wurden die Untersetzer noch etwas grober abgefeilt und der Draht wurde verdreht. So sollte der Eindruck erweckt werden, dass die Artefakte von Drittklässlern hergestellt wurden.

Anschließend wurden die Untersetzer von mir mithilfe eines Kriterienkatalogs bewertet. Die einzelnen Kriterien wurden bewertet und es wurden jeweils die erreichten Prozentwerte eingetragen. Die Prozentwerte wurden addiert und der Mittelwert errechnet. So ergab sich für jedes Artefakt eine Note. Auch hier wurden zwei verschiedene Leistungsniveaus festgestellt. Die Notenfindung erfolgte auf Basis eines selbst erstellten Notenschlüssels (s. Tab. 1), der sich an verschiedenen Notenschlüsseln für die Grundschule orientierte. Da es für Baden-Württemberg keinen einheitlichen Notenschlüssel für die Grundschule gibt, sondern dieser von Schule zu Schule und von Fach zu Fach unterschiedlich ausfallen kann, wurde sich für einen linearen Notenschlüssel entschieden, bei dem mit ca. 50% der Leistungserwartung eine glatte 4 erreicht wird. Die ganzen Noten in der Tabelle beinhalten auch die Viertelnoten. Um diese zu bestimmen, werden die Prozentpunkte gleichmäßig aufgeteilt. Bei der Bewertung mithilfe des Kriterienkatalogs wurde sich für eine Punktevergabe in Prozent entschieden, damit das Ergebnis so genau wie möglich ist und nicht durch Entscheidungen beim Runden von Punkten verfälscht wird.

Auch die vorherige Bewertung mit Hilfe des Kriterienkatalogs stellte zwei verschiedenen Leistungsniveaus fest: Artefakte A und B wurden beide mit einer 1,75, Artefakt C mit einer 3,25 und Artefakt D mit einer 3,00 benotet.

Bei der Fertigungsaufgabe liegt der Fokus auf der Herstellung eines Artefaktes, inklusive der Bedienung der Werkzeuge, dem Verwenden von Fachbegriffen, der Sachkompetenz, der Einhaltung der Sicherheitsvorschriften durch die Lernenden und auf dem Endprodukt selbst. Anhand der gewählten Artefakte kann beobachtet werden, ob die Lernenden Anweisungen befolgen können, ob sie gegebene Längen und Abstände messen können und ob sie mit den benötigten Werkzeugen umgehen können. Ebenso kann beobachtet werden, inwieweit die Lernenden auf Ästhetik achten und wie sauber sie arbeiten. Die Probanden hatten also eine große Auswahl an Faktoren, anhand derer sie ihre Benotung durchführen konnten.

#### Namen zur Operationalisierung des Migrationshintergrunds

In einer Untersuchung zu Namen und ihrer vermeintlichen Herkunft innerhalb einer Studie zur Bewertung von Schüler\*innenarbeiten im Fach Deutsch haben Bonefeld und Dickhäuser (2018) festhalten können, dass der Name Murat laut 98,1 % der 52 Probanden eindeutig türkischer Herkunft ist und der Name Max eindeutig als deutscher Name (98,1 %) identifiziert wird. Außerdem konnten sie feststellen, dass alle türkische Namen von den Probanden generell mit einer geringeren Intelligenz verknüpft wurden als die als deutschstämmig empfundenen Namen. Die Namen Max und Murat unterscheiden sich in ihrer Bewertung bezüglich Intelligenz und Attraktivität am wenigsten, weshalb sich hier – wie Bonefeld und Dickhäuser – für diese beiden Namen entschieden wurde. In einer Befragung von N = 12 (50% weiblich; M = 34.83 Jahre alt, SD = 1.88) Probanden, bei der sich an dem Vorgehen von Bonefeld und Dickhäuser orientiert wurde, sollten die Namen Eymen, Yüksel, Ömer, Simon, Jonas und Justus nach Herkunft (Türkisch; Deutsch; keine Angabe), Intelligenz (gar nicht intelligent = 1; sehr intelligent = 5) und sozialem Status (sehr niedrig = 1; sehr hoch = 5) bewertet werden. Alle Namen wurden mit 100% als eindeutig türkische bzw. deutsche Namen bewertet. Den geringsten Unterschied bezüglich der Intelligenz hatten Ömer (M = 3.25, SD = 0.7) und Simon (M = 3.42, SD = 0.49). Beide wurden ebenfalls ähnlich in Bezug auf ihren sozialen Status eingeordnet (Ömer, M = 2.58, SD = 0.86; Simon, M = 2.75, SD = 0.83). Anhand dieser Untersuchung wurde der Name Ömer für den türkischen Jungen und der Name Simon für den deutschen Jungen festgelegt.

Note	1	1,5	2	2,5	3	3,5	4
Prozent (%)	100-92	91-87	86-77	76-72	71-62	61-57	46-47

Tabella 1: Angewendeter Notenschlüssel, orientiert an Notenschlüssel für die Grundschule Baden-Württemberg

## HYPOTHESEN

Mit Hilfe der Untersuchung sollte überprüft werden, ob (1) die Leistungsbewertung von Lehramtsstudierenden in Abhängigkeit des Migrationshintergrundes variiert und ob (2) die gleiche Leistung bei hergestellten Topfuntersetzern von Schülern mit angeblichem türkischem Hintergrund schlechter benotet wird, als bei Schülern mit vermeintlich deutschem Hintergrund. Abhängige Variable war der türkische Migrationshintergrund der Schüler. Unabhängige Variablen waren zum einen die experimentelle Bedingung (gute vs. durchschnittliche Leistung), zum anderen die Noten, die in Baden-Württemberg zur Leistungsbewertung zur Verfügung stehen. Die empirischen Hypothesen lauten daher:

- H1: Es wird erwartet, dass statistisch signifikante Abweichungen zwischen den Bewertungen der Arbeiten deutscher und türkischer Schüler, unabhängig vom tatsächlichen Leistungsniveau, auftreten.
- H2: Die Arbeiten der türkischen Schüler werden im Schnitt schlechter benotet, als dieselben Arbeiten der deutschen Schüler.

Um eine möglichst hohe Teststärke zu erhalten, sollten so viele Studierende wie möglich befragt werden. Die Bedingungen, von welchen die Stichprobengröße abhing, war zum einen die begrenzte Zeit und zum anderen die begrenzte Bereitschaft der Studierenden, an der Studie teilzunehmen. Somit lagen gegen Ende der Durchführungsphase Daten von  $N = 44$  Lehramtsstudenten (100% weiblich) im Alter zwischen 19 und 42 Jahren ( $M = 22.61$  Jahre,  $SD = 3.58$ ) vor, die an der Pädagogischen Hochschule in Schwäbisch Gmünd studierten. Sie waren zum Zeitpunkt der Interviews im Bachelor- oder Masterstudiengang für das Grundschullehramt eingeschrieben ( $M = 4.82$  Semester,  $SD = 1.79$ ;) und studierten als zweites Fach den Schwerpunkt Sachunterricht. Die Teilnehmer wurden nach ihrem Studiengang ausgewählt, entweder durch direkte Ansprache oder über ihre Dozenten.

## STUDIENAUFBAU

Die Studie hatte ein  $2 \times 2$  faktorielles experimentelles Design. Den Teilnehmerinnen wurden vier verschiedene Unter-setzer aus Holz präsentiert, von denen zwei jeweils eine gute bzw. eine durchschnittliche Leistung darstellten (Faktor: Leistungsniveau). Durch Zuhilfenahme eines selbsterstellten Kriterienkatalogs wurden die guten Leistungen (Artefakte A und B) und die durchschnittlichen Leistungen (Artefakte C und D) im Vorfeld benotet. Die Unter-setzer wurden vermeintlich von männlichen, achtjährigen Schülern einer dritten Klasse erstellt, von denen zwei offensichtlich einen Migrationshintergrund hatten und zwei offensichtlich keinen (Faktor: Migrationshintergrund). In den ersten Interviews wurden die guten Leistungen den beiden immigrierten Schülern zugeschrieben und die durchschnittlichen Arbeiten den nicht-immigrierten. In den späteren Interviews wurden die Leistungen jeweils den anderen Schülern zugeschrieben. Das Leistungsniveau wurde in den letzten Interviews noch

einmal vertauscht, um eine gleiche Anzahl an Bewertungen beider Leistungsniveaus zu erhalten. Die Teilnehmerinnen wurden einzeln befragt. Sie saßen an einem Tisch, auf dem die vier Artefakte der Schüler lagen. Sie wurden darüber informiert, dass sie an einer Studie zur Benotung technischer Artefakte im Sachunterricht teilnahmen. Zunächst machten sie Angaben zu ihrem Alter, ihrem Geschlecht, ihrem Studiengang und zu dem Semester, in welchem sie sich befanden. Danach wurden ihnen die Arbeiten der Schüler präsentiert, wobei auch die Namen, das Alter und die Klassenstufe der einzelnen Schüler genannt wurden. Dann wurden ihnen die einzelnen Durchführungsschritte bei der Herstellung genannt. Daraufhin sollten sie die Artefakte unter Anwendung des baden-württembergischen Notensystems (Viertelnoten im Bereich von 1,00 bis 6,00 mit 1,00 als beste und 6,00 als schlechteste mögliche Leistung) benoten und ihre Wahl begründen. Nach Nennung einer Note für ein Artefakt, wurden sie gebeten den Namen des eben bewerteten Schülers zu nennen, wodurch ihre Angaben validiert werden sollten. Anschließend sollten sie Kriterien nennen, die sie bei ihrer Notenfindung nicht beachtet haben oder nicht beachtet haben konnten, die man außerdem zur Benotung von Artefakten aus dem technischen Sachunterricht heranziehen könnte. Diese letzte Frage wurde in die Befragung mit aufgenommen, um aus ihren Antworten auf mögliche Bewertungsverfahren für technische Artefakte in der Grundschule zu schließen. Falls sich keine Unterschiede in der Benotung technischer Artefakte von Lernenden mit und ohne Migrationshintergrund zeigen sollten, kann aus der Studie zumindest ein Kriterienkatalog erstellt werden, den man den Studierenden an die Hand geben kann. So hätten sie für ihre Tätigkeit als Lehrkraft einen Rahmen, nach dem sie Artefakte ihrer zukünftigen Lernenden bewerten können.

Nach Abschluss des Interviews konnten sie Informationen darüber erhalten, wie Studierende mit Schwerpunkt Technik die Artefakte bewertet hätten und wie sie dabei vorgegangen wären. Häufig folgte darauf noch ein kurzes Gespräch über Nachteile des Studiums des Faches Sachunterricht, da dieses Fach im Studium in einzelne Fächer unterteilt ist. Die häufigste Befürchtung der Studierenden ist das quasi fachfremde Unterrichten des Sachunterrichts im Lehrberuf, da in der sachunterrichtlichen Ausbildung ein Großteil der zu unterrichtenden Themen nicht gelehrt wird. Neben dem Hinterfragen dieses Ausbildungsaspektes wurde als weiterer Aspekt bemängelt, dass keine Übungen zur Notengebung angeboten werden; ein weiterer großer Teil des angestrebten Lehrberufs.

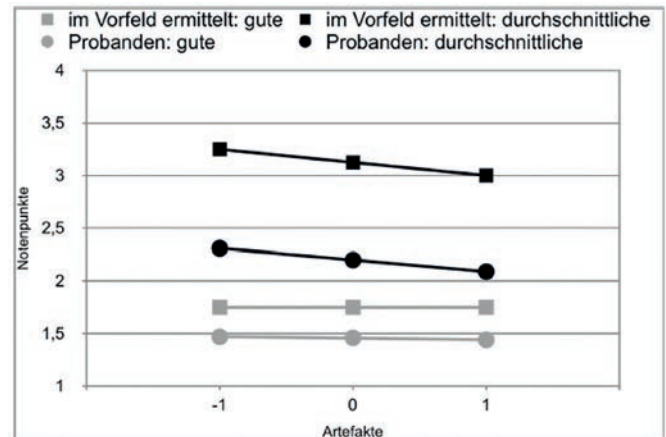
Die Teilnehmerinnen hatten keine Informationen zum sozioökonomischen Status der Schüler. Dieser konnte nur durch die gewählten Namen der vermeintlichen Schüler hergeleitet werden. Aus diesem Grund wurden die Namen so gewählt, dass sie einer eindeutigen nationalen Herkunft zugeordnet werden konnten (vgl. Bonefeld & Dickhäuser 2018, S. 5).

## ERGEBNISSE

## Vorbereitung der Daten

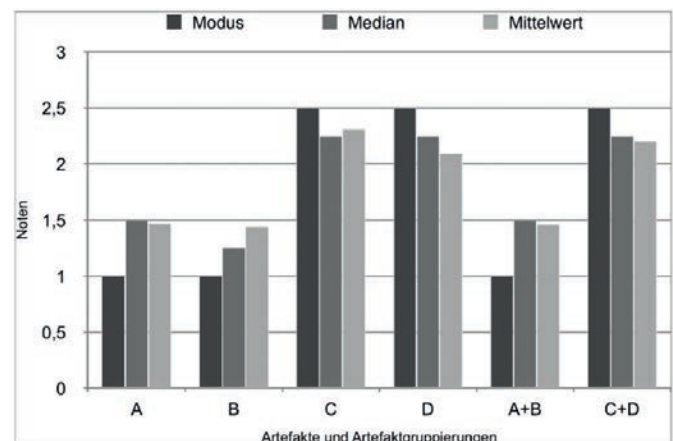
Die durchgeführten Interviews wurden im Nachhinein transkribiert. Die von den Probandinnen  $N = 44$  vergebenen Noten wurden notiert und es wurden verschiedene Durchschnittsnoten ermittelt. Außerdem wurden die vergebenen Noten in eine Rangfolge gebracht, da die Noten objektiv sind und es große Spannen zwischen den Noten verschiedener Probandinnen gab. Auf diese Weise konnte so die von den Probanden herangezogene soziale Bezugsnorm abgebildet werden. Die beste Note, die von einer Probandin vergeben wurde, belegte in der Rangfolge also Platz 1 und die schlechteste Note derselben Probandin belegte Platz 4. Wurden gleiche Noten vergeben, so wurden die theoretischen Ränge aufgeteilt und beide Artefakte bekamen denselben halben Rang. Belegten zwei Artefakte also den zweiten Platz, wurde beiden Artefakten der Rang 2,5 zugeteilt. So konnten trotz der Unterschiede in der Notenvergabe und trotz einigen Ausreißern, also gültigen Werten, die stark vom mittleren 50%-Bereich abweichen, alle Bewertungen in die Auswertung aufgenommen werden. Zunächst wurden die Durchschnittsnoten für die einzelnen Artefakte ( $M_A - M_D$ ), für die guten Artefakte insgesamt ( $M_{A+B}$ ) und für die durchschnittlichen Artefakte insgesamt ( $M_{C+D}$ ) unabhängig von den vermeintlichen Schülern ermittelt, so dass daran die Abweichung der Durchschnittsnote für den deutschen bzw. türkischen Lernenden für die einzelnen Artefakte bzw. Artefaktgruppierungen (gut, durchschnittlich) abgeleitet werden kann. Danach wurden die Durchschnittsnoten der einzelnen Schüler (Murat:  $M_{TM}$ , Ömer:  $M_{TÖ}$ , Max:  $M_{DM}$ , Simon:  $M_{DS}$ ) berechnet und diese mit den durchschnittlichen Bewertungen für die Artefakte verglichen. Ebenso wurde die Leistung der türkischen ( $M_T$ ) und deutschen ( $M_D$ ) Schüler insgesamt berechnet, um diese miteinander zu vergleichen. Die Standardabweichung ( $S_p$ ) wurde für die Mittelwerte berechnet, um die Streuung feststellen zu können. Um die Effektstärke ( $d$ ) zu messen, wurde auf die Effektstärke nach Cohen zurückgegriffen. Da der Mittelwert von Ausreißern beeinflusst wird, wurden außerdem bei den Noten und den Platzierungen der Modalwert ( $\bar{x}_d$ ) und Median ( $\bar{x}$ ) berechnet. Der Modalwert wird auch Modus genannt und nennt den Wert, der am häufigsten genannt wurde. Der Median teilt eine geordnete Reihe der Beobachtungsdaten in zwei Hälften, wodurch Ausreißer keinen Einfluss auf ihn haben.

In Abbildung 1 kann gesehen werden, dass die guten Artefakte (Artefakte A und B) von den Probandinnen insgesamt als die besseren Artefakte bestimmt und auch dementsprechend benotet wurden: Artefakt A hat eine Durchschnittsnote von  $M_A = 1.47$ ,  $SD = 0.5$  erreicht, Artefakt B von  $M_B = 1.44$ ,  $SD = 0.4$ . Diese Artefakte wurden im Vorfeld mithilfe des Kriterienkatalogs mit jeweils der Note 1,75 bewertet. Damit weichen die Bewertungen der Probandinnen um ca. eine Viertelnote in positiver Richtung ab. Die guten Artefakte insgesamt erhielten somit eine Durchschnittsbewertung von  $M_{A+B} = 1.46$ ,  $SD = 0.45$ . Im Vorfeld wurden sie mit der Note 1,75 im Durchschnitt ermittelt. Außerdem kann in Abbildung 1 abgelesen werden, dass die durchschnittlichen Artefakte (Artefakte C und D) insgesamt von den Probanden als die schlechteren Artefakte erkannt wurden: sie erhielten



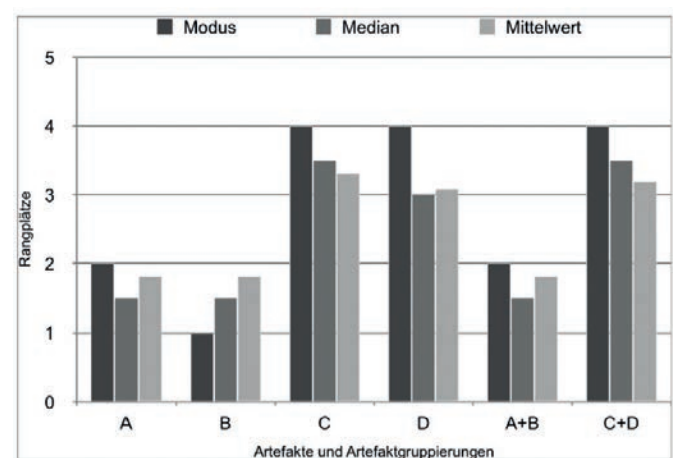
grau: - 1 =  $M_A$ , 0 =  $M_{A+B}$ , 1 =  $M_B$ ; schwarz: - 1 =  $M_C$ , 0 =  $M_{C+D}$ , 1 =  $M_D$

Abbildung 1: Durchschnittliche Notenvergabe für die guten und durchschnittlichen Artefakte im Vorfeld und von den Probandinnen



Artefakte A+B = gute Leistungen, Artefakte C+D = durchschnittliche Leistungen

Abbildung 2: Modus, Median und Mittelwert der von den Probandinnen indirekt vergebenen Platzierungen für die einzelnen Artefakte und Artefaktgruppierungen



Artefakte A+B = gute Leistungen, Artefakte C+D = durchschnittliche Leistungen

Abbildung 3: Modus, Median und Mittelwert der von den Probandinnen indirekt vergebenen Platzierungen für die einzelnen Artefakte und Artefaktgruppierungen



Durchschnittsnoten von  $M_C = 2.31$ ,  $SD = 0.5$  bzw.  $M_D = 2.09$ ,  $SD = 0.56$ . Im Vorfeld wurde Artefakt C mit der Note 3,25 und Artefakt D mit der Note 3,00 bewertet. Auch hier bewerteten die Probandinnen beide Artefakte besser als die im Vorfeld festgelegte Bewertung: jeweils um ca. 1,00 Noten. Die durchschnittlichen Artefakte insgesamt wurden also von den Probanden mit  $M_{C+D} = 2.20$ ,  $SD = 0.53$  und im Vorfeld mit einem Durchschnitt von 3,125 bewertet.

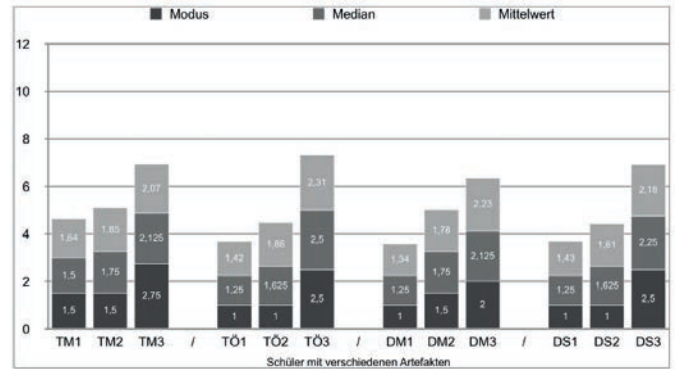
Abbildung 2 zeigt, dass bei den Artefakten A und B am häufigsten (jeweils 15-mal) die Note 1,00 vergeben wurde ( $\bar{x}_{dA} = 1.00$ ;  $\bar{x}_{dB} = 1.00$ ). Der Median von Artefakt A liegt bei  $\tilde{x}_A = 1.50$  und von Artefakt B bei  $\tilde{x}_B = 1.25$ . Insgesamt wurde Artefakt B also mit den besten Noten bewertet. Dies spiegelt sich auch in der Rangfolge ( $M_R$ ) wider (Abb. 3): Artefakt B liegt 13-mal, und somit am häufigsten von allen Artefakten, auf Platz 1 ( $\bar{x}_{dRB} = 1.0$ ) und hat einen Median von  $\tilde{x}_{RB} = 1.50$ . Durchschnittlich liegt es in der Rangfolge auf Platz  $M_{RB} = 1.81$ ,  $SD = 0.74$ . Artefakt A hat einen Modus von  $\bar{x}_{dRA} = 2.00$  (14 Nennungen). Der Median liegt bei  $\tilde{x}_{RA} = 1.50$ . Durchschnittlich liegt Artefakt A wie Artefakt B auf Platz  $M_{RA} = 1.81$ , hat aber eine größere Streuung ( $SD = 0.79$ ).

An Artefakt C wurden die Noten 2,00 und 2,50 am häufigsten (jeweils 9-mal) vergeben ( $\bar{x}_{dC} = 2.00, 2.50$ ). Der Median in der Notenvergabe liegt bei  $\tilde{x}_C = 2.25$ . Somit nimmt Artefakt C am häufigsten (18-mal) Platz 4 ( $\bar{x}_{dRC} = 4.00$ ) in der Rangfolge ein und hat hier einen Median von  $\tilde{x}_{RC} = 3.50$ . Im Durchschnitt belegt Artefakt C den Rang  $M_{RC} = 3.31$ ,  $SD = 0.62$  und ist somit das von den Probandinnen am schlechtesten bewertete Artefakt. Dies deckt sich mit der im Vorfeld vorgenommenen Bewertung der Artefakte. Artefakt D bekam eine Durchschnittsnote von  $M_D = 2.09$ ,  $SD = 0.56$ . Der Modus ist  $\bar{x}_{dD} = 2.50$  mit acht Nennungen und der Median liegt wie bei Artefakt C bei  $\tilde{x}_D = 2.25$ . In der Rangfolge unterscheidet sich Artefakt D allerdings leicht von Artefakt C: es hat auch den Modus  $\bar{x}_{dRD} = 4.00$ , allerdings mit nur 14 Nennungen. Der Median ist hier  $\tilde{x}_{RD} = 3.00$  und die durchschnittliche Platzierung ist  $M_{RD} = 3.08$ ,  $SD = 0.87$ .

Werden die guten Artefakte (A+B) zusammengenommen, wurden sie 30-mal mit der Note 1,00 benotet ( $\bar{x}_{dA+B} = 1.00$ ). Sie haben einen Median von  $\tilde{x}_{A+B} = 1.50$ . Wird die Rangfolge betrachtet, nehmen sie im Mittel Platz  $M_{RA+B} = 1.81$ ,  $SD = 0.76$  ein. Hier haben sie den Modus  $\bar{x}_{dRA+B} = 2.00$  (25 Nennungen) und einen Median von  $\tilde{x}_{RA+B} = 1.5$ . Für die durchschnittlichen Artefakte (C+D) zusammengenommen, wurde ein Modus von  $\bar{x}_{dC+D} = 2.50$  mit 17 Nennungen berechnet. Der Median liegt bei  $\tilde{x}_{C+D} = 2.25$ . Am häufigsten belegen die durchschnittlichen Artefakte Platz  $M_{RC+D} = 3.19$ ,  $SD = 0.83$ . Hier haben sie einen Modus von  $\bar{x}_{dRC+D} = 4.00$  (32 Nennungen) und einen Median von  $\tilde{x}_{RC+D} = 3.5$ .

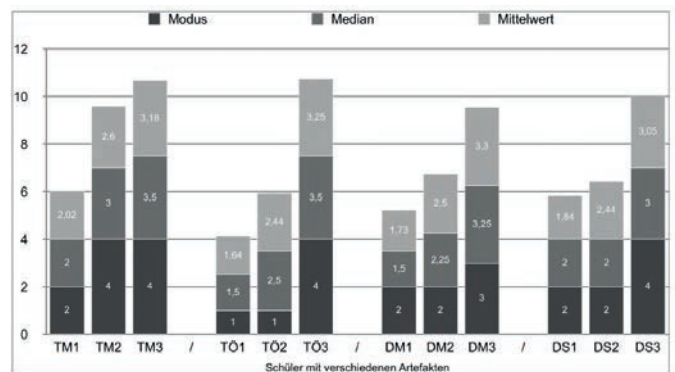
### SCHÜLER

Der Name Murat hat von den Probanden eine Durchschnittsnote von  $M_{TM} = 1.85$ ,  $SD = 0.67$  bekommen (für Notenverteilung s. Abb. 4). Der Median liegt bei  $\tilde{x}_{TM} = 1.75$  und der Modus ist  $\bar{x}_d = 1.50$  mit 11 Nennungen. Im Durchschnitt wurde Murat auf Rang  $M_{RTM} = 2,60$ ,  $SD = 1.07$  platziert (für Rangverteilung s. Abb. 5). Hier liegt der Median bei  $\tilde{x}_{RTM} = 3.00$  und



TM: türkisch Murat; TÖ: türkisch Ömer; DM: deutsch Max; DS: deutsch Simon; 1 = gute Artefakte; 2 = alle Artefakte; 3 = durchschnittliche Artefakte

Abbildung 4: Modus, Median und Mittelwert der von den Probandinnen vergebenen Noten für die Artefakte nach einzelnen Schülern



TM: türkisch Murat; TÖ: türkisch Ömer; DM: deutsch Max; DS: deutsch Simon; 1 = gute Artefakte; 2 = alle Artefakte; 3 = durchschnittliche Artefakte

Abbildung 5: Modus, Median und Mittelwert der von den Probandinnen indirekt vergebenen Platzierungen für die einzelnen Artefakte und Artefaktgruppierungen

der Modus bei  $\bar{x}_d = 4.00$  mit 10 Nennungen. Ömer hat eine Durchschnittsnote von  $M_{TÖ} = 1.86$ ,  $SD = 0.76$  bekommen. Der Median liegt hier bei  $\tilde{x}_{TÖ} = 1,625$  und der Modus ist  $\bar{x}_d = 1.00$  mit 10 Nennungen. Ömer wurde somit indirekt durchschnittlich auf Rang  $M_{RTÖ} = 2,44$ ,  $SD = 1.17$  gesehen. Der Median ist  $\tilde{x}_{RTÖ} = 2.50$  und der Modus ist  $\bar{x}_d = 1.00$  mit 11 Nennungen. Somit haben die türkischen Schüler eine Durchschnittsnote von  $M_T = 1,86$ ,  $SD = 0.69$  bekommen. Der Median ist  $\tilde{x}_T = 1.75$  und der Modus ist  $\bar{x}_d = 1.50$  mit 19 Nennungen. In der Rangfolge liegen sie durchschnittlich auf Platz  $M_{RT} = 2,52$ ,  $SD = 1.11$ . Der Median liegt bei  $\tilde{x}_{RT} = 1.75$  und der Modus ist  $\bar{x}_d = 4$  mit 18 Nennungen.

Der Name Max bekam die Durchschnittsnote  $M_{DM} = 1,78$ ,  $SD = 0.62$ ; der Median ist  $\tilde{x}_{DM} = 1.75$  und der Modus ist  $\bar{x}_d = 1.00, 2.00$  mit jeweils 9 Nennungen. Seine Noten konnten im Durchschnitt Platz  $M_{RDM} = 2,50$ ,  $SD = 1.01$  erreichen, mit einem Median von  $\tilde{x}_{RDM} = 2.25$  und einem Modus von  $\bar{x}_d = 2.00$  mit 10 Nennungen. Der Name Simon hat die Durchschnittsnote  $M_{DS} = 1,81$ ,  $SD = 0.78$  erhalten. Hier liegt der Median bei  $\tilde{x}_{DS} = 1.625$  und der Modus bei  $\bar{x}_d = 1.00$  mit 12 Nennungen. Seine Artefakte lagen im Schnitt auf Platz  $M_{RDS} = 2,44$ ,  $SD = 1.02$ . Der Median liegt bei  $\tilde{x}_{RDS} = 2.00$ , ebenso wie

der Modus  $\bar{x}_D = 2.00$  mit 11 Nennungen. Somit erhielten die deutschen Schüler im Durchschnitt eine Note von  $M_D = 1,79$ ,  $SD = 0.70$ . Der Median ist  $\tilde{x}_D = 1.75$  und der Modus liegt bei  $\bar{x}_D = 1.00$  mit 21 Nennungen. Der Durchschnittliche Rang beträgt  $M_{RD} = 2,47$ ,  $SD = 1.01$ . Der Median liegt bei  $\tilde{x}_{RD} = 2.00$ , ebenso wie der Modus  $\bar{x}_{RD} = 2.00$  mit 21 Nennungen.

Die Notengebung für die guten Artefakte ist relativ gleich, allerdings sind die Noten für Murat mit einem Abstand am schlechtesten. In der Rangfolge schneidet Ömer am besten ab: er belegte am häufigsten Platz 1, während die anderen Schüler sich den zweiten Platz nach Nennungen teilen. Bei den schlechten Artefakten schneiden die türkischen Schüler in der Notenverteilung insgesamt schlechter ab, als die deutschen Schüler. Das spiegelt sich auch in der Rangfolge wider. Hier belegt Max insgesamt die besten Plätze. Besonders deutlich wird hier, dass Ömer den größten Unterschied in der Notenvergabe und der Rangbelegung zwischen den guten und den schlechten Artefakten hat. Allerdings gab es bei der Rangfolge auch Unterschiede zu verzeichnen: 45,5 % der Probanden haben türkische Schüler bei der Bewertung der guten Artefakte nicht auf die Plätze 1 oder 2 gesetzt, sondern auf einen schlechteren Platz. Bei den deutschen Schülern taten dies 22,7 % der Probanden. Ebenso wurden die türkischen Schüler bei der Bewertung der durchschnittlichen Artefakte von 36,4 % der Probanden die Leistung besser platziert, als auf die Plätze 3 oder 4. Als den deutschen Schülern die durchschnittlichen Artefakte zugeteilt waren, haben 45,5 % der Probanden ihre Leistung besser bewertet als die gute Leistung der türkischen Schüler.

### Hypothesentests

Um Hypothese (1) zu testen, wurden die durchschnittlichen Artefaktbewertungen mit den durchschnittlichen Bewertungen der türkischen bzw. deutschen Schüler verglichen. Ebenso wurden die zugeteilten Ränge verglichen und auf Auffälligkeiten hin untersucht. Insgesamt erhielten die türkischen Schüler eine Durchschnittsnote von  $M_T = 1,86$  und die deutschen Schüler von  $M_D = 1,80$ . Hier kann nur ein sehr geringer Effekt ( $d = 0.09$ ) in Bezug auf den Migrationshintergrund festgestellt werden. Ähnlich verhält es sich bei den Noten für die besseren Artefakte ( $d = 0.05$ ). Lediglich bei den schlechteren Artefakten hat der Migrationshintergrund Auswirkungen auf die Note. Bei einem  $d = 0.33$  kann ein geringer Effekt festgestellt werden: die türkischen Schüler werden

bei den durchschnittlichen Artefakten noch schlechter bewertet als die deutschen Schüler.

Um Hypothese (2) zu testen, wurden die Leistungen der türkischen Schüler für jedes Artefakt mit den Leistungen der deutschen Schüler verglichen. Abbildung 6 zeigt die festgestellten Notendurchschnitte der Artefakte, als sie türkischen Schülern zugeteilt waren im Vergleich zu den Notendurchschnitten der deutschen Schüler. Es wird ersichtlich, dass es einen geringen Effekt ( $d = 0.33$ ) gibt, wenn türkische und deutsche Schüler nach den guten Artefakten beurteilt werden. Bei den durchschnittlichen Artefakten gibt es allerdings keinen Effekt ( $d = -0.03$ ) des Migrationshintergrundes auf die Leistungsbewertung. Durch die unregelmäßig verteilten Rangfolgen der Artefakte kann in der Vergabe der Ränge kein Effekt ( $d = 0.06$ ) festgestellt werden.

Neben den zuvor festgelegten Hypothesen konnte festgestellt werden, dass der Name Murat insgesamt am schlechtesten bewertet wurde und dies mit deutlichem Abstand. War ihm ein gutes Artefakt zugeteilt, belegte er trotzdem von den Probanden indirekt zugeteilt zweimal den 4. Rang in der Rangfolge; er wurde also trotz seines guten Artefaktes zweimal von allen Schülern am schlechtesten beurteilt. Hier wurden ihm die Noten 2,25 und 2,5 gegeben, was ja zunächst nicht schlecht erscheint. Da aber alle Probanden die soziale Bezugsnorm zur Notenfindung herangezogen haben, zeigt dies deutlich, dass hier der Migrationshintergrund eine Rolle spielte.

Als weitere Feststellung kann festgehalten werden, dass die Probanden Probleme hatten, Noten festzulegen und sich über die Ausbildung diesbezüglich beschwerten.

### DISKUSSION

Es wurde untersucht, ob es einen Einfluss des Migrationshintergrundes eines Lernenden auf die Vergabe von Noten gibt. Dazu bewerteten die Probandinnen vier Topfunter-setzer aus Holz, die zwei verschiedenen Leistungsniveaus aufwiesen. Es wurde davon ausgegangen, dass Schüler mit Migrationshintergrund von den Probandinnen benachteiligt werden würden. Da hier die Probandinnen nicht auf ein regelbasiertes Urteil zurückgreifen konnten, sondern sie sich die Kriterien, nach denen sie benoten, selbst überlegten, kann von der Notenvergabe auf eventuelle Vorurteile den migrierten Schülern gegenüber geschlossen werden. In zu-

Leistungsniveau	Note			Rang		
	M (SD)		Cohen's d	M (SD)		Cohen's d
	türkische Schüler	deutsche Schüler		türkische Schüler	deutsche Schüler	
gut	1.53 (0.46)	1.39 (0.40)	0.33	1.83 (0.86)	1.78 (0.65)	0.06
durchschnittlich	2.19 (0.56)	2.21 (0.50)	-0.03	3.22 (0.88)	3.17 (0.62)	0.06
Cohen's d	-1.09	-1.43		-1.60	-2.17	
insgesamt	1.86 (0.69)	1.79 (0.70)	0.09	2.52 (1.11)	2.47 (1.01)	0.04

Abbildung 6: Durchschnittliche Werte und Standardabweichung (in Klammern) der Noten und Rangplätze getrennt nach Leistungsniveau und Migrationshintergrund

künftigen Studien könnte untersucht werden, ob es einen Unterschied macht, wenn ein regelbasiertes Verfahren die Grundlage der Notenfindung bildet. So könnten eventuell größere Effekte aufgedeckt werden, ähnlich wie es in den Studien von Bonefeld et. al. (2017; 2018) der Fall war. Dennoch konnte beobachtet werden, dass es einen geringen Effekt auf die Notengebung hat, wenn der Schüler einen Migrationshintergrund hat. Im Gesamten kann gesagt werden, dass Schüler mit Migrationshintergrund bei gleicher Leistung schlechter benotet werden als deutsche Schüler. Dies wurde besonders bei dem Namen Murat deutlich. Er wurde bei den guten und den durchschnittlichen Artefakten am schlechtesten beurteilt. Auch wenn der Name Ömer im Vorfeld im Rahmen einer kleinen Untersuchung als eindeutig türkischer Name identifiziert worden ist, kann es sein, dass die Studierenden mit dem Namen nichts verbinden konnten, weshalb die Notengebung bei ihm ganz anders ausfiel als bei Murat. In zukünftigen Untersuchungen kann überprüft werden, ob es mit einem anderen türkischen Namen zu anderen Ergebnissen kommt.

Es konnte außerdem beobachtet werden, dass alle Probandinnen die soziale Bezugsnorm zur Notenfindung zu Rate zogen. Dies macht eventuell Sinn, da es sich nur um vier Artefakte handelte, wodurch es sich anbot, sie miteinander zu vergleichen. Ebenso konnte beobachtet werden, dass die Probandinnen Probleme bei der Notenfindung hatten. Sie konnten keine einheitlichen Kriterien nennen, auf die sie ihre Notenfindung stützen. Viele Probandinnen entschieden nach der Optik. Der technische Aspekt wurde nur selten und dann auch nicht im Ganzen, sondern nur einzelne Teilgebiete genannt. Dies deckt sich mit den Aussagen der Probanden, dass sie mit der Ausbildung zur Lehrkraft für den Sach-

unterricht diesbezüglich unzufrieden sind. Sie bemängeln, dass sie nicht allumfassend für diesen ausgebildet werden, sondern nur einen Teilbereich als Schwerpunkt studieren. Ebenso bemängelten sie, dass sie nicht lernen, wie Noten vergeben werden. Gerade in Bezug auf technische Artefakte und ihre Bewertung äußerten sie große Bedenken.

#### Autorinneninformation

Sarah Schüssler-Hanenberg  
B.ed.



studiert an der PH Schwäbisch Gmünd Grundschul-lehramt mit den Fächern Deutsch und Technik. Sie bedankt sich bei allen Studentinnen, die an der Studie teilgenommen und sie unterstützt haben. Sie appelliert an alle Studierenden, im Fall einer zukünftigen Befragung durch Kommilitonen an dieser teilzunehmen.

#### Literatur

- Bonefeld, M., Dickhäuser, O., Janke, S., Praetorius, A.-K. & Dresel, M. (2017). Migrationsbedingte Disparitäten in der Notenvergabe nach dem Übergang auf das Gymnasium. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 49 (1), 11-23
- Bonefeld, M. & Dickhäuser, O. (2018). (Biased) Grading of Students' Performance: Students' Names, Performance Level, and Implicit Attitudes. *Frontiers in Psychology*, 9:481
- Darley, J. M. & Gross, P. H. (1983). A hypothesis-conforming bias in labeling effects. *Journal of Personality and Social Psychology*, 44, 590-598
- Glock, S., Krolak-Schwerdt, S. & Pit-ten Cate, I.M. (2015). Are school placement recommendations accurate? The effect of students' ethnicity on teachers' judgments and recognition memory. *Eur J Psychol Educ* 30, 169-188
- Jussim, L. & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: knowns and unknowns, resolved and unresolved controversies. *Pers. Soc. Psychol.* 9 (2), 131-155
- Kahraman, B. & Knoblich, G. (2000). Stechen statt Sprechen. Valenz und Aktivierbarkeit von Stereotypen über Türken. *Zeitschrift für Sozialpsychologie* 31, 31-43
- Nickerson, R. S. (1998). Confirmation Bias. A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology* 2, 175-220
- Walton, G. M. & Spencer, S. J. (2009). Latent ability: grades and test scores systematically underestimate the intellectual ability of negatively stereotyped students. *Psychological Science*, 20 (9), 1132-1139
- Weidlich, J. & Spannagel, C. (2014). Die Vorbereitungsphase im Flipped Classroom. Vorlesungsvideos versus Aufgaben. In K. Rummel (Hrsg.), *Lernräume gestalten – Bildungskontexte vielfältig denken* (Medien in der Wissenschaft, 67, S. 363-367). Münster: Waxmann. <https://doi.org/10.25656/01:10105>.
- Wing, J. M. (2006). Computational Thinking“. *Communications of the ACM*, 49(3), 33-35. <https://doi.org/10.1145/1118178.1118215>.