

Gendersensibilität als Aspekt von Unterrichtsqualität in den Fächern Deutsch und Mathematik?

ALFRED LINDL, MARA RADER UND LAURA SIMBÖCK

1 Forschungshintergrund und -anlass

1.1 Genderspezifische Leistungsdisparitäten in schulischen Kompetenzbereichen

Seit einem Vierteljahrhundert dokumentieren die Ergebnisse (inter-)nationaler Schuleistungsstudien wie der Bildungstrend des Instituts zur Qualitätsentwicklung im Bildungswesen (IQB), das Programme for International Student Assessment (PISA), die Trends in International Mathematics and Science Study (TIMSS) oder die Internationale Grundschul-Lese-Untersuchung (IGLU) regelmäßig geschlechtsbezogene Leistungsunterschiede im deutschen Bildungssystem (z. B. Baumert et al., 2001; Bos et al., 2003; Kölner et al., 2010; Mullis et al., 2000). Diese erweisen sich über die verschiedenen Vergleichsstudien hinweg als relativ stabil (z. B. Diedrich et al., 2023; Frey et al., 2023; Heine et al., 2023; Nonte et al., 2020) und sind deren Fokus gemäß vor allem für Deutsch und Mathematik untersucht, wobei Mädchen in der Regel bessere Leistungen im Lesen, Jungen in Mathematik erreichen (z. B. Gentrup et al., 2023; Lewalter et al., 2023). Diese genderspezifischen Befunde stellen sich spätestens gegen Ende der Primarstufe ein (z. B. Lorenz et al., 2023; Nonte et al., 2020) und können sich während der Sekundarstufe I weiter ausprägen (z. B. Stanat et al., 2018).

Die Erklärungsansätze hierzu sind vielfältig und berücksichtigen je nach Forschungsschwerpunkt unterschiedliche Faktoren wie biologische, psychosoziale oder soziologische Merkmale (zsf. Rieggle-Crumb et al., 2018; Stanat et al., 2018; Wang & Yu, 2023). In der Bildungswissenschaft werden hierfür unter anderem unterrichtliche, schulische oder institutionelle Aspekte als zentrale Einflussvariablen für Lernerfolge, aber auch die Motivations- und Interessensentwicklung von Schülerinnen und Schülern in Betracht gezogen (Encinas-Martín & Cherian, 2023; Stanat et al., 2018). So finden sich Hinweise darauf, dass sich professionelle Überzeugungen von Lehrkräften (z. B. Einstellungen/Erwartungshaltungen hinsichtlich der Geschlechter) wie auch schülerseitige Merkmale (z. B. Selbstkonzept, Stereotypwahrnehmung) in den Deutsch- und Mathematikleistungen niederschlagen (z. B. Copur-Gencturk et al., 2023; Jansen et al., 2019; Muntoni et al., 2019; Olczyk et al., 2023; Salikutluk & Heyne, 2017). Wenig untersucht ist bislang jedoch die Rolle des Unterrichts, also inwiefern in dessen Ge-

staltung geschlechtsbezogene Überzeugungen von Lehrkräften erkennbar werden, wie diese darin mit Annahmen zu genderspezifischen Leistungsunterschieden umgehen und ob diese durch professionelles Lehrkraftverhalten moderiert werden können.

1.2 Gendersensibilität als Aspekt von Unterrichtsqualität

Nach Berliner (2005) gilt Unterricht als qualitätsvoll, wenn er normativen Standards entspricht und nachweislich lernwirksam ist. Lehr-Lernsituationen sollten grundsätzlich die gesellschaftlich anerkannte und gesetzlich verankerte Chancengleichheit für Schülerinnen und Schüler gewährleisten (KMK, 2016). Dies allein ist – wie obige Befunde verdeutlichen (Abschnitt 1.1) – allerdings noch keineswegs hinreichend, um genderspezifische Leistungsdisparitäten zu egalisieren bzw. gar nicht erst entstehen zu lassen. Um also bezüglich Gendersensibilität qualitätsvollen Unterricht zu erreichen, bedarf es zusätzlich einer bewussten Gestaltung des unterrichtlichen Angebots durch die Lehrkraft (Helmke, 2022). Diese beschränkt sich nicht auf Oberflächenmerkmale wie Sozialformen oder Methodenvielfalt, die als weniger bedeutsam für Lernerfolge gelten, sondern betrifft tiefenstrukturelle Prozessmerkmale wie kognitive Aktivierung oder konstruktive Unterstützung, welche sich als besonders lernförderlich erwiesen haben (z. B. Decristan et al., 2020; Kunter et al., 2011; Seidel & Shavelson, 2007).

Beispielsweise berichten Hochweber und Vieluf (2018) auf Basis von Daten aus der DESI-Studie (Deutsch Englisch Schülerleistungen International) positive Zusammenhänge zwischen hohen Ausprägungen in den Prozessmerkmalen effektives Klassenmanagement, angemessenes Lerntempo sowie kognitive Aktivierung und geringeren Geschlechtsdifferenzen bezüglich der Entwicklung der Leseleistung. Auch hinsichtlich der Lesemotivation reduzierte ein hohes Maß an Unterstützung und kognitiver Aktivierung genderspezifische Unterschiede. Dabei kann die Wahrnehmung dieser Qualitätskriterien jedoch selbst vom Geschlecht abhängen, wie Jurik et al. (2015) sowie Stang und McElvany (2021) für das Fach Deutsch belegen, in dem Mädchen – sogar unter Kontrolle der Lesekompetenzwerte – zu positiveren Einschätzungen neigen als Jungen; für Mathematik fanden sich bei Jurik et al. (2015) hingegen keine Beurteilungsunterschiede. Während diese Befunde das Potenzial eines qualitätsvollen Unterrichtsangebots zum Ausgleich von Leistungsdifferenzen zwischen Jungen und Mädchen andeuten, fehlen bislang empirische Untersuchungen, inwiefern hierfür eine gendersensible Unterrichtsgestaltung förderlich ist. An diesem Punkt setzt der vorliegende Beitrag an.

1.3 Fragestellungen

Die nachfolgenden Studien entstanden im Kontext der interdisziplinären Forschungsgruppe FALKO-PV (Fachspezifische Lehrkraftkompetenzen – Prädiktive Validierung)¹ an der Universität Regensburg. Ihr Forschungsfokus liegt auf der Untersuchung der Zusammenhänge zwischen Fach- und fachdidaktischem Wissen von Lehrkräften, Qualitätsmerkmalen von Unterricht sowie dessen Zielkriterien in sechs Schulfächern,

¹ Die Nachwuchsforschungsgruppe FALKO-PV wird im Rahmenprogramm Empirische Bildungsforschung für fünf Jahre (2021–2026) aus Mitteln des Bundesministeriums für Bildung und Forschung (BMBF) unter dem Förderkennzeichen 01G2103 gefördert. Für diese Veröffentlichung sind die Autorinnen und der Autor verantwortlich.

darunter Deutsch und Mathematik. In projektinternen Vorüberlegungen zu unterrichtlichen Qualitätsmerkmalen wurde aufgrund des einleitend vorgestellten Forschungsstands (Abschnitt 1.1) der Aspekt der Gendersensibilität in beiden Fächern diskutiert und als relevant erachtet. Daraus ergaben sich verschiedene Fragestellungen, die im Zentrum dieses Beitrags stehen:

1. (Wie) Lässt sich die Gendersensibilität von Deutsch- und Mathematikunterricht durch Schülerinnen und Schüler kriterienbasiert einschätzen?
2. Welche Zusammenhänge bestehen zwischen Gendersensibilität und anderen Merkmalen von Unterrichtsqualität?

Eine Untersuchung dieser Fragen verspricht Einblicke in die gendersensible Gestaltung von Deutsch- und Mathematikunterricht sowie Erkenntnisse zur Bedeutung von Gendersensibilität als Qualitätsmerkmal. Daraus lassen sich Anregungen für die Unterrichtsentwicklung und Lehrkräftebildung ableiten.

2 Design und Methode

2.1 Stichprobe und Durchführung

Die Gesamtstichprobe setzt sich aus drei Studien zusammen. An der ersten, einer Präilotierung, die im Frühsommer 2022 als Onlinefragebogen durchgeführt wurde, nahmen Schülerinnen und Schüler der Sekundarstufe I in Deutsch ($N = 15$ aus 2 Klassen) und Mathematik ($N = 71$ aus 6 Klassen) teil. An der fortlaufenden Pilotierung, die über die Evaluationsanwendung AMADEUS (anonym nutzbare mobile App zur digitalen Evaluation des Unterrichts durch Schüler:innen) erfolgt, beteiligten sich bislang 135 Schülerinnen und Schüler in Deutsch (aus 11 Klassen) und 271 in Mathematik (aus 16 Klassen). Aus Gründen der Datensparsamkeit und Anonymisierung wurden keine Kovariablen (z. B. Geschlecht, Alter) erfasst.

In die noch andauernde Hauptstudie sind derzeit 122 Schülerinnen und Schüler (57 % weiblich) aus fünf sechsten Klassen in Deutsch (bei fünf Lehrerinnen) und 169 Schülerinnen und Schüler (57 % weiblich) aus acht achten Klassen in Mathematik (bei fünf Lehrern und einer Lehrerin) an bayerischen Gymnasien involviert. Bei allen drei Studien handelt es sich um Gelegenheitsstichproben, die über persönliche Kontakte, Einladung via E-Mail oder Eigeninitiative der Lehrkraft gewonnen wurden. Die Mitwirkung aller Teilnehmenden erfolgt freiwillig, unentgeltlich und unter Wahrung datenschutzrechtlicher und forschungsethischer Bestimmungen.

2.2 Konzeptualisierung und Operationalisierung von Gendersensibilität

Um Gendersensibilität mittels kriterienbasierter Einschätzungen von Schülerinnen und Schülern zu erfassen (zu Beurteilungen aus Schülersicht z. B. Kleickmann et al., 2019; Kunter & Baumert, 2006; Lindl et al., 2024), wurde eine Skala mit sechs Items entwickelt, von denen zwei positiv (Nr. 2, 3) und vier invers formuliert sind (Nr. 1, 4, 5, 6; vgl. Tabelle 1). Zur besseren Vergleichbarkeit waren die Stimuli für Deutsch und Mathe-

matik bis auf die Fachnennung identisch. Bei den Items 1 und 6 wurde zudem jeweils dasjenige Geschlecht eingesetzt, welches stereotyp als begabter gilt (z. B. Hand et al., 2017; Muntoni & Retelsdorf, 2018; Abschnitt 1.1). Inhaltlich wird Gendersensibilität somit über divergierende Potenzialzuschreibungen, Leistungserwartungen sowie Fördermaßnahmen abgebildet, wie sie durch die Lehrkraft im Unterricht zum Ausdruck gebracht werden können. Eine vollständige Übersicht über die Items bietet Tabelle 1.

Tabelle 1: Items zu Gendersensibilität im Deutsch- und Mathematikunterricht

Item-nummer	Itemtext Deutsch	Item-nummer	Itemtext Mathematik
1a ^{inv./*}	Ich habe das Gefühl, dass unsere Lehrkraft <i>Mädchen in Deutsch</i> mehr zutraut.	1b ^{inv./*}	Ich habe das Gefühl, dass unsere Lehrkraft <i>Jungen in Mathematik</i> mehr zutraut.
2a	Unsere Lehrkraft vermittelt mir den Eindruck, dass jeder in <i>Deutsch</i> gut sein kann.	2b	Unsere Lehrkraft vermittelt mir den Eindruck, dass jeder in <i>Mathematik</i> gut sein kann.
3a*	Ich habe das Gefühl, unsere Lehrkraft ist davon überzeugt, dass Leistungen in <i>Deutsch</i> nicht vom Geschlecht abhängen.	3b*	Ich habe das Gefühl, unsere Lehrkraft ist davon überzeugt, dass Leistungen in <i>Mathematik</i> nicht vom Geschlecht abhängen.
4a ^{inv.}	Unsere Lehrkraft fördert Jungen im Unterricht besonders.	4b ^{inv.}	Unsere Lehrkraft fördert Jungen im Unterricht besonders.
5a ^{inv.}	Unsere Lehrkraft fördert Mädchen im Unterricht besonders.	5b ^{inv.}	Unsere Lehrkraft fördert Mädchen im Unterricht besonders.
6a ^{inv./*}	Unsere Lehrkraft vermittelt den Eindruck, dass <i>Mädchen Deutsch</i> einfach besser können als <i>Jungen</i> .	6b ^{inv./*}	Unsere Lehrkraft vermittelt den Eindruck, dass <i>Jungen Mathematik</i> einfach besser können als <i>Mädchen</i> .

Bem.: Als Antwortskala diente eine fünfstufige Likert-Skala von 0 = Trifft nicht zu, ..., 4 = Trifft zu. *inv.* kennzeichnet invers formulierte Items, * diejenigen Items, die in Studie 2 Verwendung finden.

2.3 Implementation und Instrumente

Die Items der Skala Gendersensibilität wurden zur Präpilotierung in einen Onlinefragebogen zu Unterrichtsqualitätsmerkmalen integriert, dessen Konzeptualisierung auf dem Syntheseframework von Praetorius et al. (2020a) und dessen sieben Dimensionen fußt. Dieser wurde zunächst als LimeSurvey-Umfrage in Deutsch und Mathematik am Ende des Schuljahrs 2021/2022 eingesetzt. Auf Basis der daraus resultierenden Erkenntnisse wurde die Webapp AMADEUS entwickelt und in zwei Versionen bereitgestellt: In ihrer frei zugänglichen Onlinevariante (amadeus.falko-pv.de) sind die drei inhaltlich und psychometrisch repräsentativsten der obigen sechs Items enthalten (Nr. 1, 3, 6). Die in der Hauptstudie eingesetzte Vollversion enthält alle sechs Items. Zur Beantwortung dient in allen Instrumenten eine fünfstufige Likert-Skala von 0 (Trifft nicht zu) bis 4 (Trifft zu).

3 Ergebnisse

3.1 Deskriptive Ergebnisse im Fächervergleich

Um zu überprüfen, inwiefern die neu konstruierten Items sensitiv für eine gendersensible Gestaltung von Deutsch- und Mathematikunterricht sind, wurden zunächst die mittleren Einschätzungen auf Itemebene betrachtet. Wie aus Tabelle 2 für das Fach Deutsch hervorgeht, entspricht in den drei Stichproben die mittlere Antworttendenz bei allen Items ihrer Formulierungsrichtung, sodass inverse Items (Nr. 1a, 4a, 5a, 6a) geringere Zustimmungswerte erhielten. Zugleich weist jedes Item auf Klassenebene eine bestimmte Spannweite zwischen niedrigster (M_{Min}) und höchster durchschnittlicher Beurteilung (M_{Max}) auf (z. B. Nr. 1a, 3a oder 6a in der Pilotierung bzw. Hauptstudie, Tabelle 2). Dies impliziert nicht nur, dass zwischen den Klassen eine gewisse Varianz in der wahrgenommenen Gestaltung gendersensiblen Unterrichts vorliegt, sondern diese mittels dieser Items – zumindest zum Teil – auch erfasst werden kann. Darauf deutet außerdem hin, dass in der Pilotierung zwei von drei und in der Haupterhebung die niedrigsten Mittelwerte aller (rekodierten) Items jeweils zu einer bestimmten Klasse gehören. Bei den höchsten mittleren Ausprägungen gilt dies wiederum für zwei von drei Items in der Pilotierung bzw. für vier von sechs Items in der Haupterhebung.

Die interne Konsistenz der Skala Gendersensibilität im Deutschunterricht wurde pro Studie mittels des Reliabilitätsindikators Cronbachs Alpha bestimmt. Während sie in der Präpilotierung angesichts der geringen Stichprobe nur $\alpha = 0,32$ beträgt, ist sie in der Pilotierung mit $\alpha = 0,79$ bzw. in der Hauptstudie mit $\alpha = 0,80$ angemessen (Tabelle 2).

Tabelle 2: Deskriptive Itemkennwerte im Fach Deutsch

Itemnr.	Präpilotierung ($N = 15$, 2 Klassen)				Pilotierung ($N = 135$, 11 Klassen)				Hauptstudie ($N = 100$, 5 Klassen)			
	M	SD	M_{Min}	M_{Max}	M	SD	M_{Min}	M_{Max}	M	SD	M_{Min}	M_{Max}
1a ^{inv.}	1,67	0,98	1,57	2,00	1,18	1,45	0,30	3,06	1,48	1,30	1,05	2,08
2a	3,27	1,16	3,17	3,71	---	---	---	---	2,96	1,04	2,17	3,13
3a	2,80	1,26	2,83	3,29	3,32	1,18	1,17	3,85	3,14	1,21	2,33	3,50
4a ^{inv.}	2,00	1,25	2,00	2,29	---	---	---	---	1,19	1,32	0,59	1,75
5a ^{inv.}	1,80	1,32	1,29	2,50	---	---	---	---	1,38	1,33	0,95	2,92
6a ^{inv.}	1,33	1,54	1,00	1,67	0,77	1,19	0,11	2,41	1,09	1,33	0,52	2,25
Skala	2,54	0,61	2,31	2,81	3,20	1,05	1,53	3,83	2,83	0,89	1,92	3,19
$\alpha = 0,32$				$\alpha = 0,79$				$\alpha = 0,80$				

Bem.: M Mittelwert, SD Standardabweichung, M_{Min} Minimum, M_{Max} Maximum (jeweils auf Klassenebene); α Cronbachs Alpha; Skalierung: 0 = Trifft nicht zu, ..., 4 = Trifft zu; ^{inv.} kennzeichnet invers formulierte Items. Diese wurden zum intuitiven Verständnis nicht bzw. erst bei Skalenbildung rekodiert.

Auch hinsichtlich des Mathematikunterrichts (Tabelle 3) spiegeln die mittleren Antworttendenzen die Ausrichtung der Items erwartungsgemäß wider. Die Spannweiten der Klassenmittelwerte sind hier bei einzelnen Items wie 3b bemerkenswert, da diese zwischen 2,60 und 3,80 (Präpilotierung) bzw. 2,62 und 4,00 (Pilotierung) liegen. Ebenso deuten die Unterschiede zwischen Klassenmittelwerten bei invers formulierten Items wie 1b (Präpilotierung: 0,80 bzw. 2,33; Pilotierung: 0 bzw. 1,54; Hauptstichprobe: 0,16 bzw. 1,40) und 6b (Präpilotierung: 0,47 bzw. 1,80; Pilotierung: 0 bzw. 1,71; Hauptstichprobe: 0,11 bzw. 1,21) auf differente Ausprägungen gendersensiblen Unterrichts hin. Bezieht man zudem mit ein, dass sich einerseits die niedrigsten Durchschnittswerte (M_{Min}) aller drei Items bei der Pilotierung auf dieselbe Klasse bzw. Lehrkraft beziehen und andererseits zwei der drei höchsten ebenso einer (anderen) Klasse zuzuordnen sind, lässt dies zumindest ansatzweise erahnen, wie breit das Spektrum gendersensiblen Unterrichtens in der Praxis ist, ohne dass dies generell ein gemeinsames Auftreten und eine solche systematische Verteilung aller Indikatoren implizieren muss. Die internen Konsistenzen der Skala sind in Mathematik angemessen ($\alpha \geq 0,62$; Tabelle 3).

Tabelle 3: Deskriptive Itemkennwerte im Fach Mathematik

Itemnr.	Präpilotierung ($N = 71$, 6 Klassen)				Pilotierung ($N = 271$, 16 Klassen)				Hauptstudie ($N = 143$, 8 Klassen)			
	M	SD	M_{Min}	M_{Max}	M	SD	M_{Min}	M_{Max}	M	SD	M_{Min}	M_{Max}
1b ^{inv.}	1,46	1,59	0,80	2,33	0,55	1,00	0,00	1,54	0,69	1,15	0,16	1,40
2b	2,59	1,24	1,80	3,80	---	---	---	---	3,12	1,08	2,58	3,59
3b	3,06	1,19	2,60	3,80	3,52	1,09	2,62	4,00	3,53	0,95	3,21	3,77
4b ^{inv.}	0,77	1,17	0,53	1,60	---	---	---	---	0,78	1,20	0,11	1,27
5b ^{inv.}	0,89	1,18	0,53	2,80	---	---	---	---	1,14	1,36	0,65	1,95
6b ^{inv.}	1,01	1,32	0,47	1,80	0,42	0,96	0,00	1,71	0,62	1,17	0,11	1,21
Skala	2,93	0,88	2,27	3,40	3,51	0,79	2,56	4,00	3,24	0,76	2,86	3,52
	$\alpha = 0,76$				$\alpha = 0,62$				$\alpha = 0,72$			

Bem.: M Mittelwert, SD Standardabweichung, Min Minimum, Max Maximum (jeweils auf Klassenebene); α Cronbachs Alpha; Skalierung: 0 = Trifft nicht zu, ..., 4 = Trifft zu; ^{inv.} kennzeichnet invers formulierte Items. Diese wurden zum intuitiven Verständnis nicht bzw. erst bei Skalenbildung rekodiert.

Die Ergebnisse in Mathematik decken sich somit im Wesentlichen mit denjenigen in Deutsch. Bei einem Vergleich der Skalenmittelwerte beider Fächer mithilfe unabhängiger t -Tests (mit Welch-Korrektur) zeigen sich in allen drei Studien Unterschiede kleiner bis mittlerer Effektgröße, die einzig in der Präpilotierung aufgrund der vergleichsweise geringen Stichprobe knapp nicht signifikant werden (Präpilotierung: $t(27,84) = -2,02$, $p = 0,053$, $d = 0,46$; Pilotierung: $t(211,60) = -3,08$, $p = 0,002$, $d = 0,35$; Hauptstudie: $t(192,21) = -3,72$, $p < 0,001$, $d = 0,50$). Dies deutet auf eine im Durchschnitt gendersensiblere Gestaltung von Mathematikunterricht und damit auf unterschiedliche fachspezifische Niveaus von Gendersensibilität hin, wenngleich die intra-

fachliche Spannweite von deren jeweiligen Ausprägungen mit Blick auf minimale und maximale Klassenmittelwerte erheblich ist (vgl. Tabelle 2 und Tabelle 3).

3.2 Zusammenhang von Gendersensibilität mit weiteren Kriterien von Unterrichtsqualität

Wie aus Tabelle 4 ersichtlich wird, fallen in den drei Studien und beiden Fächern alle Korrelationen zwischen dem Merkmal Gendersensibilität und den sieben Dimensionen von Unterrichtsqualität nach Praetorius et al. (2020a) erwartungskonform positiv aus. Abgesehen von drei Fällen treten substanzelle bis moderate Zusammenhänge auf (Deutsch: $0,20 \leq r \leq 0,59$; Mathematik: $0,23 \leq r \leq 0,58$), deren Größenordnungen indizieren, dass sich das Kriterium Gendersensibilität von anderen Aspekten empirisch unterscheiden lässt. Dieses ist insgesamt etwas enger mit sozio-emotionaler Unterstützung und eher weniger mit kognitiver Aktivierung assoziiert (vgl. Tabelle 4). Zudem erweisen sich die Zusammenhänge in der Pilotierung und Hauptstichprobe im Fach Deutsch größer als in Mathematik. Hier sind sie hinsichtlich der Dimensionen Auswahl und Thematisierung von Fachinhalten und -methoden ($r=0,10$), formatives Assessment ($r=0,08$) und Unterstützung des Lernens aller Schülerinnen und Schüler ($r=0,01$) in der Hauptstichprobe gering und nicht signifikant.

Tabelle 4: Interskalenkorrelationen zwischen Gendersensibilität und weiteren Kriterien von Unterrichtsqualität

Dimension	Präpilotierung (6 Items)		Pilotierung (3 Items)		Hauptstudie (6 Items)	
	Deutsch ($N = 15$)	Mathematik ($N = 71$)	Deutsch ($N = 135$)	Mathematik ($N = 271$)	Deutsch ($N = 90/$ $88/101$)	Mathematik ($N = 122/$ $122/143$)
I Auswahl und Thematisierung von Fachinhalten und -methoden	---	0,52**	0,41**	0,35**	0,31*	0,10
II Kognitive Aktivierung	---	0,29*	0,37**	0,29**	0,24*	0,23*
III Unterstützung des Übens	0,20	0,58**	0,40**	0,32**	0,44**	0,38**
IV Formatives Assessment	---	0,46**	0,44**	0,30**	0,26*	0,08
V Unterstützung des Lernens aller Schüler:innen	---	0,35*	0,42**	0,26**	0,25*	0,01
VI Sozio-emotionale Unterstützung	---	0,40*	0,56**	0,47**	0,59**	0,46**
VII Klassenführung	0,49	0,44**	0,39**	0,31**	0,40**	0,38**

Bem.: * $p < 0,05$, ** $p < 0,01$ (nach Holm, 1979, korrigiert). Aufgrund des Rotationsdesigns der Hauptstudie variiert die Stichprobengröße je nach Dimensionspaar bzw. -tripel. Die Zahlen nach N gelten somit für I, III/II, IV/V, VI, VII.

4 Diskussion und Limitationen

Vor dem Hintergrund der seit vielen Jahren dokumentierten geschlechtsbezogenen Leistungsdifferenzen in Deutsch und Mathematik und der in einzelnen Studien erkennbaren Bedeutung von Unterrichtsqualitätsmerkmalen zu deren Reduzierung versteht und modelliert dieser Beitrag Gendersensibilität als wichtigen Aspekt wirksamer Lehr- und Lernprozesse. Um empirisch greifbar zu machen, inwiefern die gendersensible Gestaltung von Unterricht zwischen Lehrkräften und möglicherweise auch Fächern variiert, wurde eine Skala mit sechs Items entwickelt und in drei Studien eingesetzt. Die daraus resultierenden Item- und Skalenkennwerte sind inhaltlich erwartungsgemäß und psychometrisch akzeptabel, sodass sich das Konstrukt Gendersensibilität mit dieser Form der Operationalisierung abbilden lässt. Die deskriptiven Ergebnisse erlauben Einblicke in den von Schülerinnen und Schülern empfundenen Status quo gendersensiblen Unterrichts in Deutsch und Mathematik. Sie verdeutlichen diesbezüglich eine nicht unerhebliche Variabilität zwischen Lehrkräften sowie ein unterschiedliches mittleres fachbezogenes Ausprägungsniveau. Eine Zusammenschau der Indikatoren legt nahe, dass der Aspekt Gendersensibilität im Unterricht zwar oftmals Berücksichtigung findet bzw. dessen Gestaltung unter diesem Kriterium in Mathematik etwas besser als in Deutsch gelingt; in beiden Fächern liegen aber auch einige Extremfälle vor, deren Unterricht das Gefahrenpotenzial birgt, genderbezogene Leistungsstereotype und -disparitäten zu festigen oder gar zu verstärken. Dies unterstreicht das Erfordernis, hierfür in der Lehrkräftebildung ein Bewusstsein zu schaffen und Gelegenheiten zur Reflexion der eigenen Einstellungen sowie Überzeugungen als zentrale Elemente der professionellen Kompetenz von Lehrkräften anzubieten (z. B. Baumert & Kunter, 2006; Blömeke et al., 2015).

Außerdem deuten Zusammenhangsanalysen mit den Merkmalen von Unterrichtsqualität nach dem Syntheseframework von Praetorius et al. (2020a) darauf hin, dass sich die Skala Gendersensibilität von diesen empirisch abgrenzen lässt und eine zusätzliche Nuance, vielleicht sogar einen weiteren fachspezifischen oder generischen Aspekt erfasst, der bislang zu „blinde[n] Flecken“ (Praetorius et al., 2020b, 303) der Unterrichtsqualitätsforschung gehört. Dem steht auch eine in einigen Interskalenkorrelationen erkennbare inhaltliche Nähe beispielsweise zu den Dimensionen Unterstützung des Übens oder sozio-emotionale Unterstützung nicht entgegen, da sich gerade in diesen – je nach unterrichtlicher Umsetzung – ein gendersensibles Angebot manifestieren kann (z. B. Übungstexte mit unterschiedlichen Geschlechtsidentitäten in Deutsch, wertschätzender Umgang mit Fehlern oder Missverständnissen ohne Potenzialzuschreibungen aufgrund des Geschlechts). Dies verweist zugleich auf konkrete Ansatzpunkte, den Aspekt Gendersensibilität bei der Planung, Durchführung und Reflexion von Unterricht mitzudenken, hierauf die Aufmerksamkeit von Lehrkräften in Ausbildung zu lenken und diesen als Gütekriterium von Lehr- und Lernprozessen zu berücksichtigen, um geschlechtsbezogenen Leistungsdifferenzen möglichst vorzubeugen.

Inwiefern eine geschlechtssensible Gestaltung hierauf tatsächlich einen direkten oder moderierenden Einfluss hat, ist Gegenstand weiterer Untersuchungen. In diese

sollten zusätzliche Kontextvariablen wie Geschlechtszugehörigkeit, Alter, genderbezogene Einstellungen der Lehrkräfte wie auch Schülerinnen und Schüler eingehen, deren Berücksichtigung aus Gründen der Datensparsamkeit und Anonymisierung hier nicht möglich war. Vor weiterreichenden Generalisierungen sollten zudem auch andere Schulfächer einbezogen werden. Trotz ihrer limitierenden Größe und Zusammensetzung erlaubt die vorliegende Stichprobe in Deutsch und Mathematik, also Fächern mit erheblichen geschlechtsbezogenen Leistungsdisparitäten, jedoch bereits einen explorativen Zugriff auf Gendersensibilität als Aspekt qualitätsvollen Unterrichts und verdeutlicht einerseits die Dringlichkeit, Genderkompetenz als wichtige Facette der Lehrkräftebildung zu begreifen und Gelegenheiten zu deren Erwerb und Förderung zu schaffen, sowie andererseits das darin enthaltene Potenzial zur fortschreitenden Unterrichtsentwicklung in Bezug auf eine Egalisierung genderspezifischer Leistungsdisparitäten und somit eine geschlechtergerechte Bildung.

Literatur

- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K. J. & Weiß, M. (Hrsg.). (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Leske + Budrich. <https://doi.org/10.1007/978-3-322-83412-6>
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9, 469–520. <https://doi.org/10.1007/s11618-006-0165-2>
- Berliner, D. C. (2005). The near impossibility of testing for teacher quality. *Journal of Teacher Education*, 56, 205–213. <https://doi.org/10.1177/0022487105275904>
- Blömeke, S., Gustafsson, J.-E. & Shavelson, R. J. (2015). Beyond Dichotomies. Competence Viewed as a Continuum. *Zeitschrift für Psychologie*, 223(1), 3–13. <https://doi.org/10.1027/2151-2604/a000194>
- Bos, W., Lankes E. M., Prenzel, M., Schwippert, K., Walther, G. & Valtin, R. (Hrsg.). (2003). *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*. Waxmann.
- Copur-Gencturk, Y., Thacker, I. & Cimpian, J. R. (2023). Teachers' race and gender biases and the moderating effects of their beliefs and dispositions. *International Journal of STEM Education*, 10, 31. <https://doi.org/10.1186/s40594-023-00420-z>
- Decristan, J., Hess, M., Holzberger, D. & Praetorius, A. K. (2020). Oberflächen- und Tiefenmerkmale. Eine Reflexion zweier prominenter Begriffe der Unterrichtsforschung. *Zeitschrift für Pädagogik. Beiheft*, 66(1), 102–116. <https://doi.org/10.3262/zpb2001102>
- Diedrich, J., Reinhold, F., Heinze, A. & Reiss, K. (2023). Mathematische Kompetenz in PISA 2022. Von Leistungsunterschieden und ihren Entwicklungen. In D. Lewalter, J. Diedrich, F. Goldhammer, O. Köller & K. Reiss (Hrsg.), *PISA 2022. Analyse der Bildungsergebnisse in Deutschland* (S. 53–86). Waxmann. <https://doi.org/10.31244/97830998488>

- Encinas-Martín, M. & Cherian, M. (2023). *Gender, Education and Skills: The Persistence of Gender Gaps in Education and Skills*. OECD Publishing. <https://doi.org/10.1787/34680dd5-en>
- Frey, A., Ludewig, U., König, C., Krampen, D., Lorenz, R. & Bos, W. (2023). Lesekompetenz von Viertklässlerinnen und Viertklässlern im internationalen Vergleich. 20-Jahre-Trend. In N. McElvany, R. Lorenz, A. Frey, F. Goldhammer, A. Schilcher & T. C. Stubbe (Hrsg.), *IGLU 2021. Lesekompetenz von Grundschulkindern im internationalen Vergleich und im Trend über 20 Jahre* (S. 111–129). Waxmann. <https://doi.org/10.31244/978383097009>
- Gentrup, S., Liebelt, A. & Lohbeck, A. (2023). Geschlechtsbezogene Disparitäten. In P. Stanat, S. Schipolowski, R. Schneider, S. Weirich, S. Henschel & K. A. Sachse (Hrsg.), *IQB-Bildungstrend 2022. Sprachliche Kompetenzen am Ende der 9. Jahrgangsstufe im dritten Ländervergleich* (S. 231–260). Waxmann. <https://doi.org/10.31244/9783830997771>
- Hand, S., Rice, L. & Greenlee, E. (2017). Exploring teachers' and students' gender role bias and students' confidence in STEM fields. *Social Psychology of Education*, 20(4), 929–945. <https://doi.org/10.1007/s11218-017-9408-8>
- Heine, J.-H., Heinle, M., Hahnel, C., Lewalter, D. & Becker-Mrotzek, M. (2023). Lesekompetenz in PISA 2022. Ergebnisse, Veränderungen und Perspektiven. In D. Lewalter, J. Diedrich, F. Goldhammer, O. Köller & K. Reiss (Hrsg.), *PISA 2022. Analyse der Bildungsergebnisse in Deutschland* (S. 139–162). Waxmann. <https://doi.org/10.31244/9783830998488>
- Helmke, A. (2022). *Unterrichtsqualität und Professionalisierung. Diagnostik von Lehr-Lern-Prozessen und evidenzbasierte Unterrichtsentwicklung*. Klett-Kallmeyer.
- Hochweber, J. & Vieluf, S. (2018). Gender differences in reading achievement and enjoyment of reading: The role of perceived teaching quality. *The Journal of Educational Research*, 111(3), 268–283. <https://doi.org/10.1080/00220671.2016.1253536>
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 65–70.
- Jansen, M., Schneider, R., Schipolowski, S. & Henschel, S. (2019). Motivationale Schülermerkmale im Fach Mathematik und in den naturwissenschaftlichen Fächern. In P. Stanat, S. Schipolowski, N. Mahler, S. Weirich & S. Henschel (Hrsg.), *IQB-Bildungstrend 2018. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I im zweiten Ländervergleich* (S. 337–354). Waxmann. <https://doi.org/10.25656/01:18131>
- Jurik, V., Häusler, J., Stubben, S. & Seidel, T. (2015). Interaction. Erste Ergebnisse einer vergleichenden Videostudie im Deutsch- und Mathematikunterricht. *Zeitschrift für Pädagogik*, 61(5), 692–711. <https://doi.org/10.25656/01:15422>
- Kleickmann, T., Praetorius, A. K. & Riecke-Baulecke, T. (2019). Beurteilung von Unterrichtsqualität. In F. Zimmermann, J. Möller & T. Riecke-Baulecke (Hrsg.), *Basiswissen Lehrerbildung. Schulische Diagnostik und Leistungsbeurteilung* (S. 207–223). Klett-Kallmeyer.

- Kultusministerkonferenz. (2016). *Leitlinien zur Sicherung der Chancengleichheit durch geschlechtersensible schulische Bildung und Erziehung* (Beschluss der Kultusministerkonferenz vom 06.10.2016/Beschluss der Konferenz der Gleichstellungs- und Frauenministerinnen und -minister, -senatorinnen und -senatoren der Länder vom 15./16.06.2016). https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluessel/2016/2016_10_06-Geschlechtersensible-schulische_Bildung.pdf
- Köller, O., Knigge, M. & Tesch, B. (Hrsg.). (2010). *Sprachliche Kompetenzen im Ländervergleich*. Waxmann.
- Kunter, M. & Baumert, J. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environments Research*, 9(3), 231–251. <https://doi.org/10.1007/s10984-006-9015-7>
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S. & Neubrand, M. (Hrsg.). (2011). *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*. Waxmann. <https://doi.org/10.31244/9783830974338>
- Lewalter, D., Diedrich, J., Goldhammer, F., Köller, O. & Reiss, K. (Hrsg.). (2023). *PISA 2022. Analyse der Bildungsergebnisse in Deutschland*. Waxmann. <https://doi.org/10.31244/9783830998488>
- Lindl, A., Lohe, V. & Kirchhoff, P. (2024). Fachspezifische Unterrichtsqualität in schulischen Fremdsprachen – Grundlagen und Forschungsansätze. In V. Lohe, A. Lindl & P. Kirchhoff (Hrsg.), *Unterrichtsqualität in schulischen Fremdsprachen* (S. 9–37). Waxmann. <https://doi.org/10.31244/978383099201.01>
- Lorenz, R., McElvany, N., Schilcher, A. & Ludewig, U. (2023). Lesekompetenz von Viertklässlerinnen und Viertklässlern im internationalen Vergleich. Testkonzeption und Ergebnisse von IGLU 2021. In N. McElvany, R. Lorenz, A. Frey, F. Goldhammer, A. Schilcher & T. C. Stubbe (Hrsg.), *IGLU 2021. Lesekompetenz von Grundschulkindern im internationalen Vergleich und im Trend über 20 Jahre* (S. 53–88). Waxmann. <https://doi.org/10.31244/9783830997009>
- Mullis, I. V., Martin, M. O., Fierros, E. G., Goldberg, A. L. & Stemler, S. E. (2000). Gender differences in achievement. IEA's Third International Mathematics and Science Study. *International Study Center, Lynch School of Education, Boston College*.
- Muntoni, F., Dunekacke, S., Heinze, A. & Retelsdorf, J. (2019). Geschlechtsspezifische Erwartungseffekte in Mathematik. Welche Rolle spielt das Professionswissen der Lehrkräfte? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 51(2), 84–96. <https://doi.org/10.1026/0049-8637/a000212>
- Muntoni, F. & Retelsdorf, J. (2018). Gender-specific teacher expectations in reading – the role of teachers' gender stereotypes. *Contemporary Educational Psychology* 54, 212–220. <https://doi.org/10.1016/j.cedpsych.2018.06.012>
- Nonte, S., Steinmayr, R. & Scholz, L. A. (2020). Geschlechterunterschiede in mathematischen und naturwissenschaftlichen Kompetenzen. In K. Schwippert, D. Kasper, O. Köller, N. McElvany, C. Selter, M. Steffensky & H. Wendt (Hrsg.), *TIMSS 2019. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 223–262). Waxmann. <https://doi.org/10.31244/9783830993193>

- Olczyk, M., Gentrup, S., Schneider, T., Volodina, A., Perinetti Casoni, V., Washbrook, E., Kwon, S. J. & Waldfogel, J. (2023). Teacher judgements and gender achievement gaps in primary education in England, Germany, and the US. *Social Science Research*, 116, 102938. <https://doi.org/10.1016/j.ssresearch.2023.102938>
- Praetorius, A. K., Herrmann, C., Gerlach, E., Zülsdorf-Kersting, M., Heinitz, B. & Nehring, A. (2020a). Unterrichtsqualität in den Fachdidaktiken im deutschsprachigen Raum – zwischen Generik und Fachspezifit. *Unterrichtswissenschaft*, 48(3), 409–446. <https://doi.org/10.1007/s42010-020-00082-8>
- Praetorius, A. K., Rogh, W. & Kleickmann, T. (2020b). Blinde Flecken des Modells der drei Basisdimensionen von Unterrichtsqualität? Das Modell im Spiegel einer internationalen Synthese von Merkmalen der Unterrichtsqualität. *Unterrichtswissenschaft*, 48(3), 303–318. <https://doi.org/10.1007/s42010-020-00072-w>
- Riegle-Crumb, C., Kyte, S. B. & Morton, K. (2018). Gender and Racial/Ethnic Differences in Educational Outcomes: Examining Patterns, Explanations, and New Directions for Research. In: B. Schneider (Hrsg.). *Handbook of the Sociology of Education in the 21st Century. Handbooks of Sociology and Social Research* (S. 131–152). Springer. https://doi.org/10.1007/978-3-319-76694-2_6
- Salikutluk, Z. & Heyne, S. (2017). Do gender roles and norms affect performance in maths? The impact of adolescents' and their peers' gender conceptions on maths grades. *European Sociological Review*, 33(3), 368–381. <https://doi.org/10.1093/esr/jcx049>
- Seidel, T. & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, 77(4), 454–499. <https://doi.org/10.3102/0034654307310317>
- Stanat, P., Bergann, S. & Taraszow, T. (Hrsg.). (2018). Geschlechtsbezogene Disparitäten im deutschen Bildungswesen. In R. Tippelt & B. Schmidt-Hertha (Hrsg.), *Handbuch Bildungsforschung* (S. 1321–1338). Springer VS. https://doi.org/10.1007/978-3-531-19981-8_59
- Stang, J. & McElvany, N. (2021). Unterschiede in der Wahrnehmung der Qualität des Deutschunterrichts zwischen Grundschülerinnen und Grundschülern. *Zeitschrift für Pädagogische Psychologie*, 35(4), 261–273. <https://doi.org/10.1024/1010-0652/a000275>
- Wang, L. H. & Yu, Z. G. (2023). Gender-moderated effects of academic self-concept on achievement, motivation, performance, and self-efficacy: A systematic review. *Frontiers in Psychology*, 14, 1136141. <https://doi.org/10.3389/fpsyg.2023.1136141>