

1 Einleitung

Mündliche Unterrichtsleistungen und ihre Bewertung machen einen Großteil des Unterrichtsgeschehens und der Zeugnisnote aus und sind damit wesentlicher und wichtiger Bestandteil von Schule. Da sich Lehrkräfte insgesamt 20-30% ihrer Arbeitszeit mit Aktivitäten befassen, die mit Leistungsbeurteilung zusammenhängen, ist eine Beschäftigung mit dieser Thematik für ihre Professionalisierung unerlässlich (vgl. Feser & Höttecke, 2020, S. 539). Die *mündliche Mitarbeit* macht dabei – je nach Bundesland, Stufe und Fach – durchschnittlich etwa 50% der jeweiligen Fachnote aus, womit sie neben den *schriftlichen Leistungen* maßgeblich die Notengebung beeinflusst (vgl. Arnold & Jürgens, 2001, S. 75). Damit wird auch ihre Bedeutung für die Lebensverläufe der Schüler:innen deutlich. Um die Möglichkeiten und Herausforderungen bei der Bewertung *mündlicher Mitarbeit* verstehen zu können, wird diese zunächst in den aktuellen Diskurs von Lernen und Leisten eingeordnet.

Relevanz des Themas

Gegenwärtig bestimmen zwei zunächst gegensätzlich erscheinende Trends die Veränderung der Lern- und Leistungskultur. Einerseits wird aufgrund gegenwärtiger Globalisierung und größerer Mobilität eine bessere Vergleichbarkeit der Abschlüsse gefordert (vgl. Niemeyer, 2022, S. 227; 229). Dafür wurden beispielsweise nationale Vergleichsarbeiten und zentrale Abschlussprüfungen sowie internationale Schulleistungstests etabliert. Die damit verbundene Fokussierung auf abfragbares kognitives Wissen und den aktuellen Lernstand lässt die individuelle Lernentwicklung in den Hintergrund rücken (vgl. Frohn & Heinrich, 2019, S. 154). Andererseits wird zunehmend die Berücksichtigung der Heterogenität der Schüler:innen gefordert, womit auch die Individualisierung des Lernens und Leistens verbunden ist (vgl. Paradies, Greving & Wester, 2018, S. 8). Diese Individualisierungstendenzen gehen auch mit veränderten Anforderungen an Schulleistungen einher. Laut Kirk (2006, S. 40) dürfen „Schulleistungen [...] nicht auf messbare Lernergebnisse reduziert werden“, da „[n]eben den kognitiven Fähigkeiten [...] Schlüsselqualifikationen und Handlungskompetenzen, d.h. übergeordnete Fähigkeiten wie z.B. Interaktionskompetenz, Teamfähigkeit, Kreativität und Fähigkeit zum selbstständigen Lernen, mit einbezogen [werden müssen]“. Da diese kaum standardisiert messbar sind, sollen sich Lehrerurteile maßgeblich durch „pädagogisch günstige Voreingenommenheit“ und weniger

durch Objektivität auszeichnen (vgl. Weinert & Schrader, 1986, S. 19). Gleichzeitig ist Objektivität vor allem für die Vergleichbarkeit von Leistungen unabdingbar, sodass die Bewertung von Leistungen zum „balancing act between those norms, expectations, rules and regulations on the one side, and teachers' beliefs and professional self-conceptualisation on the other“ (Falkenberg & Ringarp, 2022, S. 203) wird. Um die Individualisierung des Lernens und die häufig im Zuge der Vergleichbarkeit geforderte Standardisierung von Leistungen miteinander zu vereinbaren, schlagen beispielsweise Jürgens und Sacher (vgl. 2000, S. 18) eine Trennung von Lern- und Leistungssituationen vor. In Deutschland ist es jedoch gängig, sämtliche unterrichtliche Leistungen zu bewerten (vgl. Maitzen, 2012, S. 54), womit reine Lernsituationen unterbunden werden. Die unterrichtlichen Leistungen werden hierbei maßgeblich durch die sogenannte *mündliche Mitarbeit* der Schüler:innen abgebildet, welche erstaunlicherweise bisher sowohl in theoretischen als auch empirischen Arbeiten nahezu unberücksichtigt bleibt. Es stellt sich daher die Frage: Wie bewerten Lehrkräfte die *mündliche Mitarbeit*?

Die Bewertung *mündlicher Mitarbeit* in der Schule wird bereits seit spätestens der 1980er Jahre in Deutschland praktiziert (vgl. Kapitel 2.2.3). Die Einführung ihrer Bewertung resultierte dabei wesentlich aus der Hoffnung, dadurch die mündliche Beteiligung am Unterricht zu verbessern und Leistungskontrollen zeitökonomischer zu gestalten (vgl. Sach, 2017, S. 234).

Die Idee, das Mündliche zu bewerten, gab es dabei auch in anderen Ländern. So wurden in England beispielsweise bereits 1986 mündliche Prüfungen im Rahmen des GCSE-Examens durchgeführt (vgl. Thompson, 2006, S. 207). Dass dennoch die Theoriebildung und empirische Untersuchung des Mündlichen ausblieb, begründen Neuland und Peschel (2013) mit verschiedenen Trendwenden der Sprachdidaktik in den 1980er und 90er Jahren, die dazu führten, dass der Fokus auf den schriftlichen Leistungen blieb. Dazu gehören beispielsweise die Schriftlichkeitswende, die kognitive Wende und die kommunikationspsychologische Wende (vgl. ebd., S. 43). Die Hoffnung, durch die Bewertung der *mündlichen Mitarbeit* die mündliche Aktivität der Schüler:innen zu erhöhen, prägt dabei auch heute noch die Begründung ihrer Legitimation. So konnten Aziz et al. (2018) zeigen, dass die aktive Beteiligung von Schüler:innen im Unterricht zu höherer Motivation führt, ihre Lernprozesse unterstützt und ihre Kommunikationsfähigkeiten sowie Problemlösefähigkeiten verbessert (vgl. ebd., S. 211).

Warum auch neuere Forschung zur Bedeutsamkeit des mündlichen Sprachgebrauchs für den Kompetenzerwerb nicht dazu geführt hat, dass der Bewertung mündlicher Unterrichtsleistungen mehr Aufmerksamkeit gewidmet wird, bleibt ungeklärt.

Dass international der *mündlichen Mitarbeit* wenig Aufmerksamkeit geschenkt wird, ist hingegen schnell ersichtlich. So sehr es für deutsche Schüler:innen normal ist, dass ihre mündliche Beteiligung Teil der Zeugnisnote ist, so wenig selbstverständlich ist es, wenn die Leistungsbewertung im internationalen Vergleich betrachtet wird: Außer in Deutschland werden nur in sehr wenigen Ländern die mündlichen Beteiligungen der Schüler:innen zur Leistungsbeurteilung herangezogen. Wenn in anderen Ländern mündliche Kompetenzen bewertet werden sollen, dann werden hierfür zumeist mündliche Prüfungen gewählt (vgl. OECD, 2013, S. 607). Doch auch dies ist keine gängige Praxis. Eine Analyse der OECD legt dar, dass nur in sehr wenigen Ländern mündliche Fragen und Antworten Teil nationaler Prüfungen sind (Übersicht in OECD, 2013, S. 606–607). In den USA ist es dafür beispielsweise üblich, viele schriftliche Lernkontrollen zu schreiben und dafür gänzlich auf mündliche Lernkontrollen zu verzichten (vgl. Krieger, 2003, S. 77).

Ein wesentlicher Vorzug der Bewertung *mündlicher Mitarbeit* gegenüber wenigen punktuellen schriftlichen Leistungen wird darin gesehen, dass sie in jeder Stunde erfolgt und daher auch oft bewertet werden kann. Die Abhängigkeit von situativen Umständen wie beispielsweise der Tagesform der Schüler:innen wird hierdurch verringert. Außerdem haben die Schüler:innen die Möglichkeit, ihre Leistungen aus schriftlichen Prüfungen mit Hilfe der *mündlichen Mitarbeit* auszugleichen, wenn die Prüfungen unerwartet schlecht ausgefallen sind. Wenn Schüler:innen sich im Unterricht anstrengen und sich am Unterricht beteiligen, soll dies auch honoriert werden (vgl. Goerdt, 2004, S. 55–56). Gleichzeitig stellt das Ausführen kommunikativer Praktiken einen Teil der Kompetenzanforderungen dar (vgl. Abraham, 2016, S. 34), welche nur durch das Hinzuziehen mündlicher Formate adäquat erfasst werden können (vgl. Goerdt, 2004, S. 56). Und auch zum Erwerb von „high-order thinking skills and capacity to perform complex tasks“ wird der Einbezug von mündlichen Kommunikationskompetenzen als gewinnbringend angesehen (vgl. OECD, 2013, S. 642–643). Da im Zuge der neuen Lern- und Leistungskultur mehr Wert auf „authentische“ – im Sinne als

für spätere Berufe relevante – Formen der Leistungsbewertung gelegt wird, rücken auch alternative Beurteilungsformen wie mündliche Präsentationen und Diskussionen in den Fokus (vgl. ebd., S. 149), die im Zuge der mündlichen Mitarbeit bewertet werden sollen. Außerdem würde eine rein schriftliche Kompetenzmessung dem Grundsatz der proportionalen Abbildung widersprechen: Da der *mündlichen Mitarbeit* im Unterricht viel Zeit gewidmet wird, sollten die so erworbenen Kompetenzen auch abgeprüft werden (vgl. Jürgens & Sacher, 2000, S. 55–56; Kirk, 2004, S. 8).

Gleichzeitig wird die Bewertung der *mündlichen Mitarbeit* in der Praxis als besonders schwierig angesehen und erfordert von den Lehrkräften hohen Einsatz und gute diagnostische Fähigkeiten (vgl. OECD, 2013, S. 606; Arnold & Jürgens, 2001, S. 75). Im Gegensatz zu schriftlichen Leistungen, zu denen alle Schüler:innen einer Klasse zum gleichen Zeitpunkt verpflichtet werden, ist die mündliche Beteiligung (zumeist) freiwillig und erfordert daher zuallererst die Eigeninitiative der Schüler:innen. Wenn aber die mündliche Beteiligung von der Bereitschaft der Schüler:innen abhängig ist, sich klassenöffentlich zu äußern, können ohne diese Bereitschaft die erworbenen Kompetenzen gar nicht erst präsentiert werden. Damit stellt sich die Frage, inwiefern die bewertete mündliche Leistung tatsächlich Rückschlüsse auf erworbene Kompetenzen bietet (vgl. Krieger, 2003, S. 77). Die Hemmschwelle, sich mündlich zu beteiligen, dürfte vor allem deswegen sehr hoch sein, weil mündliche Beiträge sehr spontan generiert werden müssen (vgl. Abraham, 2016, S. 21) und für jede:n im Klassenraum hörbar und somit potenziell angreifbar sind. Der interaktive und reziproke Charakter mündlicher Kommunikation macht es vor allem den Schüler:innen schwer sich zu beteiligen, die beispielsweise aufgrund kognitiver und sprachlicher Defizite mehr Zeit für die Formulierung einer Antwort brauchen (vgl. Jürgens & Sacher, 2008, S. 89; Kohler, 2019, S. 12). Mit Blick auf die zunehmende sprachliche und leistungsbezogene Heterogenität an deutschen Schulen, ist diese Problematik aktueller denn je (vgl. Dotzel, 2021, S. 10). Dass nicht alle Schüler:innen in jeder Stunde und schon gar nicht zu jeder Aufgabe aufgerufen werden können, erschwert zusätzlich die Vergleichbarkeit mündlicher Beiträge (vgl. Abraham, 2016, S. 149). Die Situativität und Spontaneität mündlicher Beiträge stellen überdies auch für die Lehrkräfte eine besondere Herausforderung dar (vgl. Jung, 2013, S. 86). Mündliche Gespräche lassen sich nicht gut im Voraus planen, womit auch keine Erwartungshorizonte vorab erstellt werden können. Die retrospektive Bewertung wiederum wird durch die Flüchtigkeit des Gesprochenen erschwert

(vgl. ebd., S. 86). Der Versuch von Franz-Josef Payhuber im Jahr 1981 „Standards für die Mündlichkeit“ zu formulieren, hat sich nicht durchgesetzt, sodass es nach wie vor primär im pädagogischen Ermessen der Lehrkräfte liegt, wie mündliche Leistungen ermittelt werden (vgl. Avenarius & Hanschmann, 2019, S. 423–424). Um Willkür entgegen zu wirken, sollen Lehrkräfte in allen Phasen der Lehrerausbildung in ihrer Beurteilungskompetenz gefördert werden. So formuliert die KMK für die Bildungswissenschaften innerhalb des Kompetenzbereichs Beurteilen dezidierte Vorgaben dazu, welche Kompetenzen Lehrkräfte sowohl in den theoretischen als auch praktischen Ausbildungsabschnitten zur Bewertung von Leistungen erlernen sollen (vgl. Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2004, S. 11). Eine gesonderte Betrachtung nach einzelnen Leistungsbereichen (z.B. mündlich oder schriftlich) wird allerdings nicht vorgenommen.

Wenn Lehrkräfte mit Schüler:innen über die Bewertung der sogenannten *mündlichen Mitarbeit* sprechen, ergeben sich weitere Schwierigkeiten. Während ursprünglich neben den schriftlichen Klassenarbeiten tatsächlich vor allem die mündlichen Wortbeiträge der Schüler:innen im Plenumsgespräch zur *mündlichen Mitarbeit* gezählt werden, bestimmen – nicht zuletzt durch die Öffnung des Unterrichts und der Forderung nach Methoden- und Medienvielfalt – nun eine Reihe weiterer Aktivitäten den Unterricht (vgl. Winter, 2018, S. 8–9). Gleichzeitig wurde der schulalltägliche Sprachgebrauch kaum entsprechend angepasst. Obwohl in den schulrechtlichen Vorgaben je nach Bundesland auch sonstige oder praktische Leistungen aufgeführt werden, wird in der theoretischen Literatur den Lehrkräften vorgeworfen, keine klare Abgrenzung zwischen diesen Begriffen vorzunehmen und unter der „mündlichen Note“ alle Unterrichtsleistungen außerhalb schriftlicher Klassenarbeiten zu subsumieren (vgl. Langhammer, 1998, S. 38; Maitzen, 2012, S. 54). Empirische Überprüfungen dieser Vorwürfe gab es bisher kaum.

Die vorliegende Arbeit geht daher den Fragen nach: Welche Vorgaben gibt es zur Bewertung *mündlicher Mitarbeit*? Und wie gehen Lehrkräfte konkret bei der Bewertung *mündlicher Mitarbeit* vor?

Aufbau der Arbeit

Die Arbeit gliedert sich in einen theoretischen Teil (Kapitel 2 und 3), einen empirischen Teil (Kapitel 4 bis 8) und einen Schlussteil (Kapitel 9 und 10). Im theo-

retischen Teil werden zunächst die für das Thema *mündliche Mitarbeit* wesentlichen theoretischen Grundsteine gelegt (Kapitel 2). Dazu gehören die Definition des (schulischen) Leistungsprinzips, die Verortung des Begriffs in das Spannungsfeld zwischen Vergleichbarkeit und Individualisierung und die mit dem schulischen Leistungsprinzip verbundenen Zielsetzungen (Kapitel 2.1). Daraufhin wird untersucht, wie der in den schulrechtlichen Regelungen festgesetzte Sammelbegriff *mündliche Mitarbeit* definiert ist. Hierfür werden die internationalen und nationalen Diskurse zur Bewertung *mündlicher Mitarbeit* dargelegt (Kapitel 2.2). Kapitel 2.3 geht näher auf die Spezifika mündlicher Kommunikation ein, um die Besonderheit der Bewertung mündlicher Leistungen herausarbeiten zu können (Kapitel 2.4). Zuletzt werden in Kapitel 3 ausgewählte Befunde zur *mündlichen Mitarbeit* dargelegt. Dazu gehören sowohl Einflüsse von Umgebungsfaktoren und didaktischen Maßnahmen auf die *mündliche Mitarbeit* der Schüler:innen sowie Wirkungen, die unmittelbar durch die Anwesenheit anderer Personen bei der Erbringung mündlicher Leistungen erzeugt werden und zuletzt Einflüsse von Persönlichkeitsmerkmalen der Schüler:innen auf ihre *mündliche Mitarbeit*.

Der empirische Teil wird eingeleitet durch die Ableitung der zentralen Fragestellungen dieser Arbeit (Kapitel 4). Zuerst wird der Frage nachgegangen, welche Vorgaben und Vorschriften Lehrkräften zur Bewertung *mündlicher Mitarbeit* gemacht werden (1). Daraufhin soll eruiert werden, wie die Lehrkräfte auf Grundlage dieser Vorgaben ihre eigene Bewertung ausgestalten (2) und zuletzt inwiefern die Bewertungspraxen der Lehrkräfte Gemeinsamkeiten und Unterschiede aufweisen (3). Dafür sollen jeweils folgende Fragen geklärt werden:

- (a) Wie wird *mündliche Mitarbeit* definiert und nach welchen Kriterien soll sie bewertet werden?
- (b) Wie werden die Einzelleistungen *mündlicher Mitarbeit* dokumentiert?
- (c) Wie wird aus den Einzelleistungen ein abschließendes Urteil für die Zeugnisnote gebildet?

Den Fragestellungen gehen insgesamt drei Teilstudien nach, die im Mixed-Methods-Design angelegt sind (Kapitel 5). Zunächst werden in einer Dokumentenanalyse (Kapitel 6) schulrechtliche Vorgaben zur Leistungsmessung und -bewertung analysiert. Damit werden zweierlei Ziele verfolgt: Zum einen legen

schulrechtliche Regelungen die Rahmenbedingungen für die individuellen Bewertungspraxen der Lehrkräfte fest, womit diese nur unter Beachtung der rechtlichen Regelungen später analysiert werden können (Forschungsfrage 1). Außerdem dienen sie einer ersten explorativen Erkundung des Forschungsgegenstandes und zeigen verschiedene Ausgestaltungsmöglichkeiten auf. Letzteres begründet auch, warum nicht nur die Vorgaben für Nordrhein-Westfalen, sondern die Vorgaben aller deutschen Bundesländer betrachtet werden. Mit Hilfe der hiermit gewonnen Erkenntnisse wird dann ein Leitfaden erstellt, mit dem Fachvorsitzende aus NRW der Fächer Mathematik und Deutsch in Experteninterviews zur Bewertung *mündlicher Mitarbeit* befragt werden (Teilstudie 2, Kapitel 7). Die interviewten Fachvorsitzenden werden einerseits als Verantwortliche für schulinterne Vorgaben dazu befragt, welche Vorgaben den Lehrkräften fächerübergreifend oder fachintern zur Bewertung *mündlicher Mitarbeit* gemacht werden (Forschungsfrage 1) und andererseits als Lehrkräfte, wie sie selbst die Vorgaben individuell ausgestalten (Forschungsfrage 2). Aufgrund der Stichprobengröße können die gewonnenen Daten allerdings nur einen ersten Einblick in die individuellen Bewertungsschemata bieten und dienen damit vorrangig der Gewinnung möglicher Antwortvorgaben für geschlossene Fragen für die darauffolgende Fragebogenstudie und damit einer ersten Exploration der zweiten Forschungsfrage. In der Fragebogenerhebung, welche die dritte Teilstudie darstellt, werden Gymnasiallehrkräfte in NRW, die Mathematik oder Deutsch unterrichten, mit Hilfe eines Online-Fragebogens zu ihren Bewertungsschemata befragt (Forschungsfrage 2). Die Antworten werden außerdem dahingehend untersucht, inwiefern sich Gemeinsamkeiten und Unterschiede zwischen den Bewertungsschemata finden lassen (Forschungsfrage 3).

Im Schlussteil werden die empirisch gewonnenen Befunde zusammengeführt und vor dem Hintergrund der eingangs dargestellten Theorie diskutiert. Nachdem die Stärken und Limitationen der eigenen Untersuchungen dargelegt werden, werden Implikationen für die pädagogische Praxis und vor allem Forschungsideen, die aus dieser explorativen Studie entstehen, ausgeführt.

