

Zwischen Belastung und Bewusstsein

Eine empirische Querschnittsstudie mittels Voice-Handicap-Index (VHI) zur Erfassung der Stimmgesundheit von Lehrkräften

Anika Ebner^a, Thomas Wagner^b, Stephan Hametner^c

^aBASOP Baumgartenberg, ^bPädagogische Hochschule Oberösterreich, ^cAnton Bruckner Privatuniversität Linz
stephan.hametner@bruckneruni.at; <https://doi.org/10.17883/pa-ho-2026-02-03>

EINGEREICHT 01 JUL 2026

ÜBERARBEITET 21 AUG 2026

ANGENOMMEN 21 SEP 2026

Die Wichtigkeit der Stimme, einem zentralen Arbeitswerkzeug von Lehrpersonen, wird oft sowohl im Studium als auch im beruflichen Alltag unterschätzt. Als Folgen können situative Überlastungen bzw. chronische Stimmprobleme auftreten. Die vorliegende empirische Querschnittsstudie widmet sich der Bestandsaufnahme solcher Problemfelder. Dazu wurde bei 414 Lehrpersonen aus der Sekundarstufe der sogenannte Voice-Handicap-Index (VHI), ein international anerkannter Indikator für subjektiv wahrgenommene Stimmgesundheit, erhoben und mit potentiellen Prädiktoren in Beziehung gesetzt. Gemischte Regressionsanalysen zeigen unter anderem einen signifikanten Einfluss von Berufserfahrung, stimmlicher Vorerkrankungen und der Selbsteinschätzung der eigenen Stimme. Nach der Diskussion didaktischer Implikationen wird die Notwendigkeit, stimmpräventive Maßnahmen bereits im Lehramtsstudium aber auch im Berufsleben von Lehrpersonen stärker zu verankern, herausgearbeitet.

SCHLÜSSELWÖRTER: Stimmgesundheit, Lehrpersonen, Sekundarstufe, Voice-Handicap-Index

1. Einleitung

Die Stimme zählt zu den zentralen Arbeitswerkzeugen von Lehrkräften und ist für einen gelingenden Unterricht unverzichtbar. Dennoch zeigen internationale Untersuchungen (Roy, Merrill, Thibeault, Parsa, Gray, et al., 2004; Smith, Nichols, Lemke, Verdolini, Gray, et al., 1994), dass stimmliche Beschwerden wie Heiserkeit, Stimmüdigkeit oder funktionelle Dysphonien zu den häufigsten berufsbedingten Gesundheitsproblemen im Lehrberuf gehören.

Lehrpersonen sind aufgrund der hohen Sprechbelastung, ungünstiger Raumakustik sowie erhöhter emotionaler und sozialer Anforderungen in dieser Hinsicht besonders vulnerabel. Außerdem können stimmliche Probleme nicht nur die Unterrichtsqualität beeinträchtigen, sondern auch zu Krankenständen, Einschränkungen im Berufsalltag und langfristigen Beeinträchtigungen der Lebens-

qualität führen. Das ist in zweierlei Hinsicht bedeutsam. Einerseits zeigen Studien, dass Beschwerden häufig subjektiv wahrgenommen werden, bevor sie objektiv messbar sind. Andererseits konnte belegt werden, dass subjektive Wahrnehmung und objektive phoniatische Befunde häufig nicht übereinstimmen (Ilomäki, Lepänen, Kleemola, Tyrmi, Laukkanen, et al., 2009), wodurch die Einschätzung der Betroffenen selbst ein zentraler Aspekt der Diagnostik darstellt.

Trotz dieser hohen Relevanz des Faktors Stimme für Lehrpersonen existieren im deutschsprachigen Raum nur wenige valide empirische Untersuchungen, die die subjektive Stimmgesundheit von Lehrkräften sowie potenzielle berufliche und psychologische Einflussfaktoren systematisch erfassen. Besonders in Österreich ist dies essentiell, da systematische Datenerhebungen weitgehend fehlen.

Die vorliegende Studie versucht daher, diese Forschungslücke zu schließen, indem sie den Status quo der Stimmgesundheit österreichischer Lehrkräfte mithilfe der validierten deutschen Version des Voice-Handicap-Index [VHI] (Jacobson, Johnson, Grywalski, Silbergleit, Jacobson, et al., 1997; Nawka, Wiesmann & Gonnermann, 2003) untersucht. Durch die Kombination der VHI-Subskalen mit subjektiver Selbsteinschätzung sowie berufserfahrungs- und schulformspezifischen Variablen bietet die Untersuchung einen Einblick in potenzielle Faktoren, die zur stimmlichen Belastung beitragen. Die vorliegenden Ergebnisse liefern nicht nur wichtige Impulse für die präventive Arbeit mit der Stimme im Lehrberuf, sondern auch für die musikpädagogische und sprechwissenschaftliche Ausbildung, in der gezielte Maßnahmen zu Stimmenschutz, Wahrnehmungsschulung und Stressbewältigung zunehmend an Bedeutung gewinnen (Gaul & Lang, 2012; Ilomäki et al., 2009; Nawrocka, Garstecka, Mackiewicz-Nartowicz, Kozakiewicz-Rutkowska, Burduk, et al., 2025).

2. Theoretischer Hintergrund

Internationale Forschungen zeigen, dass Lehrpersonen aufgrund hoher Sprechbelastung, ungünstiger akustischer Bedingungen und arbeitsbedingtem Stress besonders anfällig für stimmliche Beeinträchtigungen sind. Zwischen 50 und 60 Prozent berichten im Verlauf ihrer Karriere von Stimmproblemen (Martins, Pereira, Hidalgo, & Tavares, 2014; Nusseck, Immerz, Spahn, Echternach, & Richter, 2021), wobei subjektive Beschwerden oft früher auftreten als objektive Befunde (Ilomäki et al., 2009).

2.1 Stimmgesundheit bei Lehrkräften

Stimmgesundheit ist ein latentes Konstrukt, welches in Forschungskontexten oft in Dimensionen oder Teilbereiche aufgespalten wird, so zum Beispiel in physiologische, funktionale und psychosoziale Aspekte (Jacobson et al., 1997). Ganz

allgemein bezeichnet Stimmgesundheit die Leistungsfähigkeit der Stimme in beruflichen und alltäglichen Kommunikationssituationen und ist insbesondere im Lehrberuf ein entscheidender Faktor, da diese Berufsgruppe besonders risikobelastet ist (Ubillos, Centeno, Ibañez, & Iraurgi, 2015).

Bisherige Studien zeigen, dass die Stimmgesundheit von Lehrpersonen sowohl durch berufsbiografische Faktoren als auch durch arbeitsbezogene Rahmenbedingungen beeinflusst wird. Einerseits wird zum Beispiel die Dauer der Berufstätigkeit als Indikator von Stimmbelastung diskutiert, die langfristig zu funktionellen Veränderungen und erhöhtem Beschwerderisiko führen kann (Meuret, 2017; van Houtte, Claeys, Wuyts, & van Lierde, 2011). Andererseits weisen empirische Befunde darauf hin, dass sich die stimmliche Beanspruchung je nach Schulstufe bzw. Unterrichtskontext unterschiedlich ausgeprägt sein kann, etwa aufgrund unterschiedlicher Lautstärkeanforderungen, Klassenführung oder Kommunikationsbedarf (Byeon, 2019; Munier & Kinsella, 2008).

2.2 Der VHI

Neben objektiven phoniatischen Diagnostiken gelten subjektive Erhebungsinstrumente, wie zum Beispiel der Voice-Related-Quality-of-Life-Index (Gräbel, Hoppe & Rosanowski, 2009), der Vocal-Fatigue-Index (Nanjundeswaran, Jacobson, Gartner-Schmidt & Verdolini Abbott, 2015) oder eben auch der VHI gerade bei Lehrkräften als besonders relevant, da sie subjektiv empfundene Beschwerden leicht messbar machen und psychometrischen Untersuchungen vorausgehen können.

Der VHI erhebt die subjektiv empfundene Beeinträchtigung der Stimme in drei Subskalen. Die Subskala, welche funktionale Faktoren der Stimme beschreibt, beinhaltet Aussagen, die den Einfluss der Stimmstörung einer Person auf ihre täglichen Aktivitäten beschreiben. Ein Beispielitem dafür ist „Meine Familie hat Schwierigkeiten, mich zu hören, wenn ich im Haus/in der Wohnung nach ihnen rufe“. Die Subskala, welche sich auf Emotionalität bezieht, umfasst Aussagen, welche Gefühle und emotionalen Reaktionen, die Menschen hinsichtlich ihrer Stimmprobleme haben, beschreibt. Ein Beispiel für diese Subskala ist das Item „Wegen meines Stimmproblems komme ich weniger aus mir heraus“. Die dritte Subskala, welche physische Aspekte der Stimme umfasst, beschreibt Aussagen, die die Selbstwahrnehmung von Stimmbeschwerden und Stimmmerkmalen widerspiegeln. Ein Item dieser Subskala ist beispielsweise die Aussage „Wenn ich spreche, komme ich außer Atem“.

Der VHI ist international eines der meistgenutzten Instrumente zur Erfassung vokaler Beeinträchtigung. Aufgrund der weiten Verbreitung dieser Skala in der Stimmdiagnostik (er liegt in validierten Übersetzungen für zahlreiche Sprachen vor) und seiner hohen Reliabilität und Validität gilt er in der Forschung und in der

klinischen Praxis als etabliertes Standardinstrument (Alarouj, Althekeallah, AlAli, Ebrahim, & Ebrahim, 2022); Jacobson et al., 1997; Nawka et al., 2003).

Die deutsche Version wurde von Nawka et al. (2003) validiert und zeigt eine sehr hohe interne Konsistenz, mit Cronbachs α zwischen 0,90 und 0,96, eine gute Differenzierungsfähigkeit zwischen pathologischen und gesunden Stimmen sowie vergleichbare Zusammenhänge mit subjektiver Heiserkeit wie in der Originalstudie (Jacobson et al., 1997).

Kritikpunkte in der Fachliteratur betreffen u.a. die möglichen kulturellen Unterschiede in emotionalen Reaktionsmustern, die eingeschränkte Sensitivität des VHI gegenüber schulformspezifischen Belastungen sowie die Debatte, ob die drei Subskalen immer trennscharf abgebildet werden können (Ilomäki et al., 2009; Meuret, 2017; Nawka et al., 2003; Nusseck et al., 2021).

2.3 Forschungsfragen und Hypothesen

Der vorliegende Beitrag verfolgt zwei Forschungsfragen. In der ersten Frage geht es darum, wie die Verteilung der VHI-Werte in der vorliegenden österreichischen Stichprobe ausfällt bzw. wie ausgeprägt potenzielle stimmliche Beeinträchtigungen bei österreichischen Lehrpersonen sind. Frage zwei befasst sich mit dem Einfluss potenzieller Prädiktoren auf solche stimmlichen Beeinträchtigungen.

Hinsichtlich Forschungsfrage eins vermuten wir, dass, ähnlich wie in Alarouj et al. (2022) sowie Ubillos et al. (2015), der VHI eine moderate stimmliche Belastung indizieren wird. Die Vergleichsstudie von Alarouj et al. (2022) zeigte zum Beispiel signifikant höhere VHI-Werte bei Lehrpersonen im Vergleich zur Allgemeinbevölkerung sowie einen höheren Anteil klinisch auffälliger Werte. Darüber hinaus berichteten Ubillos et al. (2015), dass 16,4 % der Lehrkräfte bereits eine Stimmstörung erlebt hatten, wobei viele keine stimmliche Ausbildung erhalten hatten. Beiden Ergebnisse verdeutlichen die erhöhte stimmliche Belastung im Lehrberuf.

Forschungsfrage zwei ist von der Hypothese geleitet, dass der VHI mit bestimmten Einflussfaktoren signifikant assoziiert ist. Unter anderem werden in Studien psychische Faktoren, stimmliche Vorerkrankungen und subjektive Selbstwahrnehmung berichtet (Meuret, 2017; Nawka et al., 2003; Nawrocka et al., 2025), aber auch schultypenspezifische sowie unterrichtliche Einflüsse (Ubillos et al., 2015).

Wir fragen daher, welche Vorhersagekraft Berufserfahrung, Schulform, Rauchverhalten, stimmliche Vorerkrankungen, die Selbsteinschätzung der eigenen Stimme bzw. Gender aufweisen.

3. Methodologie

3.1 Stichprobe

Ursprünglich wurden Daten von 541 österreichischen Lehrpersonen der Sekundarstufe online erhoben. Diese Stichprobe war weder randomisiert, stratifiziert noch vorselektiert und kann daher als convenience sample betrachtet werden. Die Erhebung, genehmigt durch die oberösterreichische Schulaufsichtsbehörde, erstreckte sich über drei Monate Mitte 2024; die Teilnahme erfolgte freiwillig und ohne finanzielle Vergütung. Alle Daten und IP-Adressen blieben anonymisiert.

Aufgrund von zu geringen Dienstzeiten, zum Teil wegen Karenzierung oder sehr geringem Beschäftigungsausmaß wurden 124 Fälle exkludiert. Daten von 417 Lehrpersonen flossen daher in die finalen Analysen ein. In dieser Stichprobe identifizierten sich 289 Lehrpersonen als weiblich, 126 als männlich und zwei als divers. Ein Drittel der Lehrpersonen unterrichtete an der Sekundarstufe 1, zwei Drittel an der Sekundarstufe 2.

3.2 Erhebungsinstrument

Als Erhebungsinstrument diente die deutsche Version¹ des ursprünglich in den USA entwickelten VHI-Fragebogens. Dieser deutsche Fragebogen elizitiert 30 Rückmeldungen zur Häufigkeit, mit der stimmbezogenen Beeinträchtigungen auftreten können. Die 30 Items bilden laut Originalstudie „funktionelle“, „physische“ sowie „emotionale“ Dimensionen mit jeweils 10 Items ab. Fünf Likert-Kategorien (0 = *nie*, 1 = *selten*, 2 = *manchmal*, 3 = *oft*, 4 = *immer*) stehen bei der Beantwortung zur Auswahl, was in individuellen Indizes zwischen 0 und 120 resultiert. Je höher der VHI, desto ausgeprägter sind stimmbezogene Beeinträchtigungen. Neben dem VHI umfasste der eingesetzte Fragebogen außerdem noch Likert-Items zu potenziellen Prädiktoren (sh. Abschnitt *Prädiktoren*).

Alle Items wurden in der browserbasierten Software *SoSciSurvey* (Leiner, 2024) als online-Befragung implementiert. Mittels eines Pilottests ($N = 30$) wurden Angaben sowie Itemformulierungen auf Verständlichkeit und Nutzerfreundlichkeit hin überprüft. Der komplette Fragebogen ist im OSF²-Repositorium abrufbar.

3.3 Prädiktoren

Der erste Prädiktor war Berufserfahrung. Diese wurde durch die Anzahl der Dienstjahre nach Huberman (1989) in vier Abstufungen operationalisiert: 1-6 Jahre, 7-18

¹ Die Validierungsstudie für den deutschsprachigen Raum von Nawka et al. (2003) mit 316 erwachsene Patienten zeigte eine hohe Gesamtreliabilität (Cronbachs $\alpha = 0,96$) der übersetzten Items.

² <https://osf.io/gkhtx>

Jahre, 19-30 Jahre und 31-40 Jahre. Der zweite Prädiktor Schulform wurde binär kodiert, mit den Ausprägungen Sekundarstufe 1 und 2. Das Rauchverhalten, Prädiktor 3, wurde durch eine viertstufige Häufigkeit abgefragt. Die Erhebung stimmlicher Vorerkrankungen waren binär kodiert (ja/nein). Genderidentifizierung war als männlich, weiblich oder divers möglich. Abschließend wurde noch die Selbsteinschätzung der eigenen Stimme als Prädiktor auf einer Likert-Skala (*normal, leicht gestört, mittelgradig gestört, hochgradig gestört*) erhoben.

3.4 Datenkodierung und Analyse

Die Datensätze enthielten keinerlei fehlende Werte (NAs), was Daten-Imputation überflüssig machte. Für die Analyse wurde die Computerumgebung R verwendet, Version 4.5.2, *[Not] Part in a Rumble* (R Core Team, 2025), insbesondere die Pakete *ltm* (Rizopoulos, 2006), *lavaan* (Rosseel, 2012) sowie *ggplot2* (Wickham, 2016).

Regressionsanalysen für die Subskalen wurden mittels linearer gemischter Modelle durchgeführt (Björk, Ruthmann, Granfors, Högväg, Andersson, 2021), basierend auf dem Paket *lme4* (Bates, Mächler, Bolker, Walker, 2015). Potenzielle Effekte der Prädiktoren wurden mittels kumulativer gemischter Modelle über die ursprünglichen Likert-Werte ohne Aggregation modelliert. Dabei kam das Paket *ordinal* (Christensen, 2022, Version 2022.11.16) zur Anwendung. Das finale Modell wurde mittels loglikelihood-Tests aufgebaut und mit der Logit-Link-Funktion und Laplace-Approximation angepasst. Es enthielt alle Prädiktoren als fixed effects sowie random intercepts für TeilnehmerInnen und Items. Haupteffekte werden in Kapitel 4 auf Grundlage von Typ-II-Wald- χ^2 -Tests berichtet, (*RVAideMemoire*, Hervé, 2022), partielle Effekte als Randeffekte (*emmeans*, Lenth, 2023, mit erweiterter Bonferroni-Methode).

4. Resultate³

4.1 Der VHI in der vorliegenden Stichprobe

Die VHI-Werte weisen mit Cronbachs $\alpha = 0,94$ (95 % CI [0,93; 0,95]) in der vorliegenden Stichprobe eine fast so hohe interne Konsistenz auf wie in der deutschen Validierungsstudie (Nawka et al., 2003). Die Subskalen zeigen ebenfalls eine solch hohe Konsistenz, mit $\alpha = 0,86$ (95 % CI [0,84; 0,88, MacDonalds $\omega = 0,91$) für „emotional“, $\alpha = 0,82$ (95 % CI [0,77; 0,83], MacDonalds $\omega = 0,87$) für „funktionell“, und $\alpha = 0,88$ (95 % CI [0,87; 0,90], MacDonalds $\omega = 0,91$) für „physisch“. Es scheint also, als ob die deutsche Version des Fragebogens das latente Konstrukt grundsätzlich valide misst.

³ Alle Rohdaten sowie das Analyse-Skript aus R sind zusammen mit diversen Zusatzmaterialien im OSF-Repository unter https://osf.io/gkhtx/overview?view_only=bd4566b8f01c4d51ba9fa46e75451de7 zu finden.

Diese drei Subskalen als Binnenstruktur des Fragebogens konnten im vorliegenden Datensatz allerdings nicht wirklich faktorenanalytisch bestätigt werden. Eine konfirmatorische Faktorenanalyse⁴ mittels diagonal gewichteter kleinste Quadrate (DWLS) zur Schätzung polychorischer Korrelationen, basierend auf einem signifikanten Bartlett-Test ($\chi^2(435) = 5549,7, p < 0,001$) sowie einem hohen Kaiser-Mayer-Olkin Kriterium (0,94), zeigte inhomogene Ladungen der Items auf die drei Subskalen. Obwohl die Modellschätzung nach 60 Iterationen erfolgreich war, und obwohl die Fit-Indizes CFI (0,99), TLI (0,98), RMSEA (0,07 (95 % CI[0,065; 0,073])) und SRMR (0,09) akzeptabel waren, passte dieses Modell nicht perfekt zu den Daten ($\chi^2(402) < 0,001$). Da die drei Subskalen jedoch in der deutschen Validierungsstudie bestätigt wurden, wurde diese Struktur nicht weiter hinterfragt bzw. neu modelliert.

Die VHI-Werte streuen in der vorliegenden Studie um den Mittelwert von $M = 44,65$, $SD = 11,96$. Die Verteilung ist allerdings ausgeprägt rechtsschief und scheint signifikant von einer Normalverteilung abzuweichen ($W[415] = 0,83, p < 0,001$). Ein Q-Q-Diagramm zeigt allerdings, dass diese Abweichungen vor allem auf sehr niedrige und hohe VHI-Werte zurückzuführen sind. Der Großteil der Werte ist somit nahezu normalverteilt und liegt vergleichsweise niedrig in der Gesamtskala. Die österreichischen Lehrpersonen weisen also im Mittel eher geringe stimmbezogenen Beeinträchtigungen auf.

Die Mittelwerte der drei Subskalen sind unterschiedlich, mit den niedrigsten Werten im Bereich „emotional“ und den höchsten für „physisch“. Ein lineares gemischtes Regressionsmodell mit random intercepts für TeilnehmerInnen und Items sowie fixed effects für die drei kontrastkodierte Subskalen ergab einen signifikanten Haupteffekt für die Subskalen (Typ-3-ANOVA mit Satterthwaite-Korrektur, $F(2;27) = 6,38, p < 0,01$). Eine Post-hoc-Analyse zeigte, dass die Subskala „emotional“ im Mittel signifikant niedriger als die drei Skalen im Durchschnitt ist ($\beta = -0,21, z = -2,72, p < 0,01$), während die Subskala „physisch“ im Mittel signifikant höher liegt ($\beta = 0,26, z = 3,37, p < 0,01$). Das bedeutet, dass stimmbezogenen Beeinträchtigungen bei österreichischen Lehrpersonen überdurchschnittlich oft im physischen Bereich zu finden sind und sich weiter weniger als emotionale Reaktionen aufgrund der Stimmproblematik äußern. Abbildung 1 illustriert die Varianz über alle 30 Items und alle 417 TeilnehmerInnen hinweg.

4 Korrelationsmatrizes, Ladungen, Standardfehler, Varianzen und Residuen sind im OSF-Repository unter https://osf.io/gkhtx/overview?view_only=bd4566b8f01c4d51ba9fa46e75451de7 zu finden.

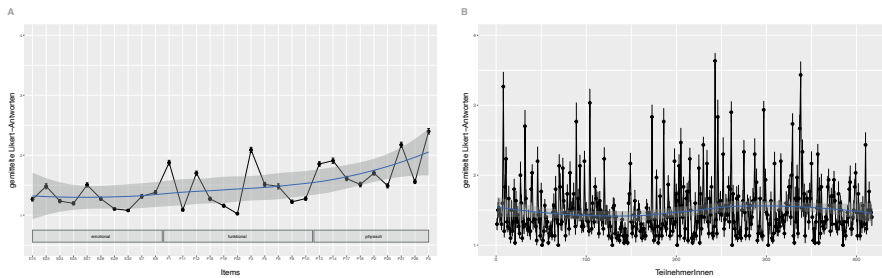


ABB. 1 Varianz der drei Subskalen

Abbildung 1, Panel A und B, zeigen eine ausgeprägte Varianz in den Likert-Antworten, sowohl bei den Items ($s^2 = 0,12$, $SD = 0,35$) als auch bei den TeilnehmerInnen ($s^2 = 0,16$, $SD = 0,41$). Während beispielsweise das Item E30 sehr niedrig bewertet wird ($M = 1,08$), erzielen die Items P4 ($M = 2,4$) und P21 ($M = 2,16$) vergleichsweise hohe Werte.

Diese Variabilität illustriert, dass die Items unterschiedliche Bewertungsmuster bei den österreichischen Lehrpersonen hervorrufen und dass sich diese in ihrer Bewertung der 30 Items erheblich unterscheiden. Die blaue Regressionslinie mit ihren Konfidenzbändern (Glättungsfunktion basierend auf t -Verteilungsapproximationen) in Panel A verdeutlicht darüber hinaus eine Steigung in der Itemvarianz von der Subskala „emotional“ links zu „physisch“ rechts. Aufgrund dieser Varianz – und auch aufgrund inhomogener Varianzen in der Prädiktorausprägungen – wurde der potenzielle Einfluss von Prädiktoren für den VHI mittels kumulativer gemischter Regressionsmodelle modelliert.

4.2 Potenzielle Effekte der Prädiktoren

Ergebnisse von Log-Likelihood-Tests legen nahe, dass das finale Modell lediglich drei signifikante Prädiktoren hat, nämlich Berufserfahrung ($\chi^2(3) = 8,10$, $p < 0,05$), stimmliche Vorerkrankungen ($\chi^2(1) = 21,59$, $p < 0,01$) und die Selbsteinschätzung der eigenen Stimme ($\chi^2(3) = 184,75$, $p < 0,012$). Das finale Modell ist in Tabelle 1 zusammengefasst.

Das kumulative gemischte Regressionsmodell zeigt eine akzeptable Modellpassung (Hessian Index = 2000, absoluter Gradient $< 0,01$). Die erklärte Varianz der abhängigen Variable liegt bei $\geq 20\%$ und scheint damit für den sozialwissenschaftlichen Bereich ausreichend.

TAB. 1 Koeffizienten und Statistiken des finalen kumulativen gemischten Modells (Laplace Approximation) für die Likert-Ratings nach Berufserfahrung, stimmlicher Vorerkrankung sowie Selbsteinschätzung der Stimme

Effekt	Schätzer		SE	z	p	Cohens d approx. ^a
	Logit	OR				
fixed effects						
Schwellenkoeffizienten						
1 2	-1,41	0,24	0,47	-3,00	< 0,001***	-0,78
2 3	1,07	2,92	0,47	2,28	< 0,005**	0,59
3 4	3,45	31,50	0,47	7,31	< 0,001***	1,90
4 5	6,03	415,72	0,50	12,05	< 0,001***	3,32
partielle Effekte für Berufserfahrung						
1 bis 6 Jahre	0,33	1,39	0,11	2,93	0,013*	0,18
07 bis 18 Jahre	-0,04	0,96	0,11	-0,40	0,689	-0,02
19 bis 30 Jahre	-0,14	0,87	0,10	-1,38	0,224	-0,08
31 bis 40 Jahre	-0,15	0,86	0,11	-1,40	0,224	-0,08
partielle Effekte für Vorerkrankung						
ja	0,26	1,30	0,19	1,36	0,173	0,14
nein	-0,26	0,77	0,19	-1,36	0,173	-0,14
partielle Effekte für Stimmeinschätzung						
hochgradig gestört	1,51	4,53	0,93	1,62	0,140	0,83
leicht gestört	-0,27	0,76	0,35	-0,79	0,432	-0,15
mittelgradig gestört	1,32	3,74	0,41	3,2	0,003**	0,73
normal	-2,55	0,08	0,33	-7,73	< 0,001***	-1,41
random intercepts	s2	SD				
TeilnehmerInnen	1,33	1,15				
Items	2,75	1,66				

Anmerkung. Anzahl der partiellen Effekte = 10, Anzahl der Beobachtungen = 12510, N = 417, OR = odds ratio, pseudo-R² (McFadden) = 0,20, Cox und Snell (ML) = 0,29.

^aCohen's d Effektstärken-Approximierung wurde mittels $\log(OR) \times \sqrt{3} / \pi$ (Sánchez-Meca, Marín-Martínez & Chacón-Moscoso, 2003) berechnet. Die Konditionsanzahl (Hessian), welche die empirische Identifizierbarkeit des Regressionsmodells misst, liegt bei 2000 und damit weit unterhalb des kritischen Bereichs. Der maximale absolute Gradient der log-likelihood-Funktion in Bezug auf die Parameter ist < 0,01 und bestätigt damit die Bedingungen für gelungene Modellkonvergenz.

Wie man in Tabelle 1 sehen kann, gibt es einen signifikanten partiellen Effekt für Berufserfahrung, nämlich für das Segment „1 bis 6 Jahre“. Die Regressionsschätzer für diese Effekte werden in den verwendeten gemischten Regressionsmodellen zunächst als Logits⁵ berechnet. Zur besseren Interpretierbarkeit sind diese allerdings auch als relative Chance bzw. Odds Ratios in Tabelle 1 angeführt.

Der Odds Ratio-Wert von 1,39 für 1 bis 6 Jahre Berufserfahrung bedeutet zum Beispiel, dass für diese Altersgruppe eine um 39 % gesteigerte relative Wahrscheinlichkeit gemessen am Durchschnitt vorliegt, höhere VHI-Werte zu bekommen, wenn alle anderen Prädiktoren im Modell konstant gehalten werden. Da höhere VHI-Werte ausgeprägtere stimmbezogene Beeinträchtigungen illustrieren, kann man daher konstatieren, dass österreichische Lehrpersonen mit der wenigsten Berufserfahrung eine signifikant höhere Wahrscheinlichkeit für stimmbezogene Beeinträchtigungen aufweisen als im Mittel. Dieser Effekt ist zwar signifikant, zeigt mit Cohens $d = 0,18$ eine jedoch nur sehr geringe Effektstärke (Cohen, 1988).

Weiterhin lassen sich aus Tabelle 1 zwei partielle Effekte für die eigene Stimm-einschätzung herauslesen („mittelgradig gestört“ und „normal“). Wie man bereits an den mathematischen Vorzeichen der beiden betroffenen Logits sehen kann, gehen diese Effekte in verschiedene Richtungen. Für alle Lehrpersonen, die ihre stimmliche Verfassung als „normal“ einschätzen, ist die relative Wahrscheinlichkeit ($OR = 0,08$), höhere VHI-Werte zu produzieren um mehr als 90 % niedriger (negativer Logit). Mit anderen Worten, diese Lehrpersonen haben sehr wahrscheinlich weniger stimmbezogene Beeinträchtigungen als alle Lehrpersonen im Mittel. Dieser Effekt ist höchst signifikant und auch mittelstark (Cohens $d = 0,73$).

Demgegenüber zeigen Lehrpersonen, welche ihre Stimme als „mittelgradig gestört“ einschätzen, eine mehr als drei Mal so hohe relative Wahrscheinlichkeit ($OR = 3,74$), höhere – also schlechtere – VHI-Werte zu produzieren wie im Mittel (positiver Logit). Und dieser Effekt ist hoch signifikant und sehr effektstark (Cohens $d > 0,8$).

Für den an sich signifikanten Haupteffekt für „stimmliche Vorerkrankungen“ lässt sich in diesem Modell kein signifikanter partieller Effekt berechnen. Der Prädiktor an sich bleibt somit ein signifikantes Element der Modellierung, seine Ausprägungen können aber nicht weiter differenziert werden.

5. Diskussion

Die Ergebnisse der vorliegenden empirischen Studie liefern einen ersten Überblick über die subjektiv wahrgenommene Stimmgesundheit von Lehrpersonen der Sekundarstufe in Österreich sowie über potenzielle Prädiktoren für eine solche Wahrnehmung.

⁵ Der Logit oder auch die logistische Einheit ist mathematisch der natürliche Logarithmus einer Chance (Odds Ratio) für eine Wahrscheinlichkeit und liegt im Bereich 0 bis 1.

Zwei Forschungsfragen wurden empirisch verfolgt. Was die Frage nach der grundsätzlichen Verteilung von VHI-Werte in der österreichischen Stichprobe angeht, so sehen wir zunächst eine hohe interne Konsistenzen sowohl für den Gesamtwert als auch für die Subskalen, was nahelegt, dass die deutsche Version des VHI das zugrundeliegende latente Konstrukt der subjektiven Stimmbeeinträchtigung valide zu erfassen scheint.

Insgesamt zeigen die Ergebnisse niedrige bis moderate VHI-Werte, was darauf hindeutet, dass österreichische Lehrpersonen im Mittel eher geringe stimmbezogene Beeinträchtigungen berichten. Gleichzeitig weist die ausgeprägte Varianz auf relevante individuelle Unterschiede innerhalb der Stichprobe hin.

Bezüglich der VHI-Dimensionen traten die höchsten Werte im physischen Bereich auf, während emotionale Belastungen signifikant niedriger ausgeprägt waren. Stimmbezogene Beeinträchtigungen betreffen somit überwiegend körperliche Aspekte wie Ermüdung oder Anstrengung und weniger emotionale Reaktionen. Dies legt nahe, dass stimmliche Belastungen im schulischen Alltag primär somatisch wahrgenommen werden. Diese Ergebnisse bestätigen Befunde von Alarouj et al. (2022) sowie Ubillos et al. (2015).

Forschungsfrage zwei befasste sich mit dem Einfluss potenzieller Prädiktoren auf stimmlichen Beeinträchtigungen. Hierzu identifizierten Regressionsanalysen drei zentrale Einflussfaktoren: Erstens zeigte sich ein Zusammenhang mit der Berufserfahrung, wobei mit zunehmenden Dienstjahren niedrigere VHI-Werte auftraten. Die höchste Wahrscheinlichkeit, stimmbezogene Beeinträchtigungen vorzuweisen, hatte die Berufsgruppe mit den wenigsten Dienstjahren. Dies könnte auf adaptive Prozesse im Berufsverlauf, ökonomischeren Stimmeneinsatz oder Bewältigungsstrategien hinweisen; alternativ ist ein Healthy-Worker-Effekt denkbar (Chowdhury, Shah & Payal, 2017).

Zweitens erwies sich die Selbsteinschätzung der eigenen Stimme als wirkmächtigster Prädiktor. Eine negative Selbsteinschätzung der eigenen stimmlichen Verfassung, vor allem in den Ausprägungen „mittelgradig“ und „hochgradig“ gestört, waren mit deutlich schlechteren VHI-Werten assoziiert. Für die Selbsteinschätzung „mittelgradig gestört“ war dieser Effekt überdies hochsignifikant und effektstark. Dieser Befund unterstreicht die zentrale Rolle subjektiver Wahrnehmungsprozesse für die Einschätzung der Stimmgesundheit und steht im Einklang mit Ergebnissen von Ilomäki et al. (2009), die zeigen, dass Selbstbewertungen der Stimme eng mit arbeitsbezogenen Faktoren und phoniatischen Befunden zusammenhängen.

Drittens zeigten stimmliche Vorerkrankungen zwar einen signifikanten Zusammenhang mit VHI-Werten im Gesamtmodell, signifikante partielle Effekte blieben jedoch aus. Dies deutet darauf hin, dass deren Einfluss durch andere Variablen überlagert werden könnte (Nawrocka et al., 2025).

Im Gegensatz zu Byeon (2019) sowie Munier und Kinsella (2008) ergaben sich in der vorliegenden Untersuchung keine signifikanten Effekte für die Schulform (Sekundarstufe I vs. II). Während diese Studien schulstufenspezifische Unterschiede in der Stimmgesundheit berichten, könnte der fehlende Effekt in der vorliegenden Stichprobe darauf hindeuten, dass institutionelle Unterschiede durch individuelle Anpassungsstrategien der Lehrpersonen kompensiert werden.

Auch für das Rauchverhalten zeigte sich kein signifikanter Effekt, was im Einklang mit Ergebnissen von Alarouj et al. (2022) steht, die ebenfalls keinen Zusammenhang zwischen Rauchen und VHI-Werten fanden. Ein möglicher Grund hierfür ist, dass der VHI subjektive Wahrnehmung erfasst, während rauchbedingte stimmliche Veränderungen nicht zwingend als Beeinträchtigung erlebt werden.

Im Gegensatz zu Befunden von Byeon (2019), die Geschlecht als signifikanten Risikofaktor für Stimmstörungen identifizieren, konnten in der vorliegenden Studie allerdings keine entsprechenden Effekte nachgewiesen werden. Dies könnte darauf hindeuten, dass geschlechtsspezifische Unterschiede durch andere Einflussfaktoren überlagert werden oder in der untersuchten Stichprobe weniger stark ausgeprägt sind.

Mit Blick auf die Praxis unterstreichen die vorliegenden Befunde insbesondere die Bedeutung früher Stimmprävention im Lehramtsstudium, die Förderung der stimmlichen Selbstwahrnehmung sowie für eine systematische Integration von Belastungs- und Stressmanagement in Aus- und Fortbildungsprogrammen. Da vor allem subjektive Faktoren eng mit der Stimmgesundheit verknüpft sind, sollten präventive Maßnahmen über rein stimmtechnische Trainingsangebote hinausgehen und gezielt weitere physische, aber auch psychosoziale Aspekte systematisch berücksichtigen.

Bisheriger Studien empfehlen bereits wiederholt ein verpflichtendes Stimmtraining in der Lehramtsausbildung, die Förderung der Selbstwahrnehmung der Stimme, Integration von Stress- und Belastungsmanagement in Aus- und Weiterbildung sowie stimmhygienische Schulungen (Ilomäki et al., 2009; Jacobson et al., 1997; Kooijman, Jong, de Thomas, Huinck, Donders, et al., 2006; Meuret, 2017; Nawka et al., 2003; Nusseck et al., 2021). Diese Erkenntnisse sind direkt auf den österreichischen Kontext übertragbar, denn gerade dort fehlen bislang institutionalisierte Präventionsprogramme.

Abschließend bleibt zu bemerken, dass die vorliegende Studie einige Limitationen aufweist und dass mehrere Einschränkungen bei der Interpretation der Ergebnisse zu berücksichtigen sind. Erstens basiert die Studie auf einer Convenience-Stichprobe, die nicht repräsentativ für alle österreichischen Lehrpersonen ist. Zweitens handelt es sich um ein Querschnittsdesign. Somit sind zum Beispiel keine zeitlich bedingten kausalen Schlussfolgerungen über Zusammenhänge zwischen Prädiktoren und Stimmgesundheit möglich. Drittens wurden ausschließlich Selbstberichtsdaten erhoben. Objektive phoniatische oder akustische Messungen

lagen nicht vor. Darüber hinaus könnte ein Selbstselektionsbias vorliegen, da besonders belastete oder besonders interessierte Personen eventuell motivierter waren, an einer solchen Studie teilzunehmen.

Grundsätzlich treffen auch die bereits im Abschnitt „Der VHI“ erwähnten Kritikpunkte in der Fachliteratur auf diese Studie zu (Ilomäki et al., 2009; Meuret, 2017; Nawka et al., 2003; Nusseck, et al., 2021). Neben potenziell problematischen psychometrischen Aspekten, wie der internen Konsistenz des Erhebungsinstruments, lassen VHI-Werte zum Beispiel keinerlei Einsichten zu potenziellen kulturellen Unterschieden in emotionalen Reaktionsmustern von Lehrpersonen zu. Außerdem hat sich gezeigt, dass schulformspezifische Belastungen mittels des VHI scheinbar nur schwer zu eruieren sind.

Zukünftige Studien sollten daher versuchen, repräsentative Stichproben zu ziehen, querschnittsartige Untersuchungen mit longitudinale Daten zu ergänzen, quantitative und qualitative Daten zu berücksichtigen und subjektive Selbsteinschätzungen mit phoniatischen oder akustische Messungen zu flankieren.

Literatur

- Alarouj, H., Althekerallah, J. M., AlAli, H., Ebrahim, M. A. & Ebrahim, M. A. K. (2022). A comparative study utilizing the Voice Handicap Index-10 (VHI-10) in teachers and the general population of Kuwait. *Journal of Voice*, 36(2), 289.e1–289.e10. <https://doi.org/10.1016/j.jvoice.2020.05.006>
- Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Björk, C., Ruthmann, S. A., Granfors, M., Högväg, J. & Andersson, S. (2021). The potential of a mixed-methods approach for research on learning to theorise music. *Music Education Research*, 23(3), 374–390. <https://doi.org/10.1080/14613808.2020.1853085>
- Byeon, H. (2019). The risk factors related to voice disorder in teachers: A systematic review and meta-analysis. *International Journal of Environmental Research and Public Health*, 16(19). <https://doi.org/10.3390/ijerph16193675>
- Chowdhury, R., Shah, D. & Payal, A. R. (2017). Healthy worker effect phenomenon: Revisited with emphasis on statistical methods – a review. *Indian Journal of Occupational and Environmental Medicine*, 21(1), 2–8. https://doi.org/10.4103/ijoom.IJOEM_53_16
- Christensen, R. H. B. (2022). *ordinal*. Regression models for ordinal data. R package (Version 2022.11.16) [Computer software]. <https://CRAN.R-project.org/package=ordinal>
- Cohen J. (1988). *Statistical power analysis for the behavioral sciences* (2. Aufl.). Lawrence Erlbaum Associates.

- Gaul, M. & Lang, S. (Hrsg.). (2012). *Voice coaching: Zum richtigen Umgang mit der Stimme im Lehrberuf*. Schneider Verlag Hohengehren.
- Gräsel, E., Hoppe, U. & Rosanowski, F. (2009). Graduierung des Voice-Related-Quality-of-Life-Index. *HNO*, 57, 896–901. <https://doi.org/10.1007/s00106-009-1953-9>
- Hervé, M. (2022). *RVAideMemoire: Testing and plotting procedures for biostatistics*. R package (Version 0.9–81-2) [Computer software]. <http://127.0.0.1:22992/library/RVAideMemoire/html/RVAideMemoire-package.html>
- Huberman, M. (1989). The professional life cycle of teachers. *Teachers College Record: The Voice of Scholarship in Education*, 91(1), 31–57. <https://doi.org/10.1177/016146818909100107>
- Ilomäki, I., Leppänen, K., Kleemola, L., Tyrmi, J., Laukkanen, A.-M. & Vilkman, E. (2009). Relationships between self-evaluations of voice and working conditions, background factors, and phoniatic findings in female teachers. *Logopedics, phoniatrics, vocology*, 34(1), 20–31. <https://doi.org/10.1080/14015430802042013>
- Jacobson, B.H., Johnson, A., Grywalski, C., Silbergleit, A., Jacobson, G., Benninger, M. S. & Newman, C. W. (1997). The Voice Handicap Index (VHI): Development and validation. *American Journal of Speech-Language Pathology*, 6(3), 66–70. <https://doi.org/10.1044/1058-0360.0603.66>
- Kooijman, P.G.C., Jong, F.I.C.R.S.de, Thomas, G., Huinck, W., Donders, R., Graa-
mans, K. & Schutte, H.K. (2006). Risk factors for voice problems in teachers. *Folia phoniatica et logopaedica: official organ of the International Association of Logopedics and Phoniatics (IALP)*, 58(3), 159–174. <https://doi.org/10.1159/000091730>
- Leiner, D. (2024). *SoSci Survey* (Version 3.5.07) [Computer software]. <https://www.soscsurvey.de/>
- Lenth, R. (2023). *emmeans: Estimated marginal means, aka least-squares means*. R package (Version 1.8.4-1) [Computer software]. <https://CRAN.R-project.org/package=emmeans>
- Martins, R.H.G., Pereira, E.R.B.N., Hidalgo, C.B. & Tavares, E.L.M. (2014). Voice disorders in teachers. A review. *Journal of Voice*, 28(6), 716–724. <https://doi.org/10.1016/j.jvoice.2014.02.008>
- Meuret, S. (2017). Welche Faktoren beeinflussen die Gesundheit der Pädagogenstimme? In M. Fuchs (Hrsg.), *Kinder- und Jugendstimme: Bd. 11. Die Stimme im pädagogischen Alltag* (S. 49–54). Logos.
- Munier, C. & Kinsella, R. (2008). The prevalence and impact of voice problems in primary school teachers. *Occupational Medicine*, 58(1), 74–76. <https://doi.org/10.1093/occmed/kqm104>
- Nanjundeswaran, C., Jacobson, B.H., Gartner-Schmidt, J. & Verdolini Abbott, K. (2015). Vocal Fatigue Index (VFI): Development and validation. *Journal of Voice*, 29(4), 433–440. doi: 10.1016/j.jvoice.2014.09.012

- Nawka, T., Wiesmann, U. & Gonnermann, U. (2003). Validierung des Voice Handicap Index (VHI) in der deutschen Fassung [Validation of the German version of the Voice Handicap Index]. *HNO*, 51(11), 921–930. <https://doi.org/10.1007/s00106-003-0909-8>
- Nawrocka, L., Garstecka, A., Mackiewicz-Nartowicz, H., Kozakiewicz-Rutkowska, A., Burduk, P. & Sinkiewicz, A. (2025). Relationship between voice impairment and stress coping styles in professionally active teachers. *Irish Journal of Medical Science*, 194(1), 333–338. <https://doi.org/10.1007/s11845-024-03816-0>
- Nusseck, M., Immerz, A., Spahn, C., Echternach, M. & Richter, B. (2021). Long-Term effects of a voice training program for teachers on vocal and mental health. *Journal of Voice*, 35(3), 438–446. <https://doi.org/10.1016/j.jvoice.2019.11.016>
- R Core Team (2025). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rizopoulos, D. (2006). *ltm*: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software*, 17(5), 1–25. <https://doi.org/10.18637/jss.v017.i05>
- Rosseel, Y. (2012). *lavaan*: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. doi:10.18637/jss.v048.i02
- Roy, N., Merrill, R. M., Thibeault, S., Parsa, R. A., Gray, S. D. & Smith, E. M. (2004). Prevalence of voice disorders in teachers and the general population. *Journal of Speech, Language, and Hearing Research*, 47(2), 281–293. [https://doi.org/10.1044/1092-4388\(2004/023\)](https://doi.org/10.1044/1092-4388(2004/023))
- Sánchez-Meca, J., Marín-Martínez, F. & Chacón-Moscoso, S. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods*, 8(4), 448–467. <https://doi.org/10.1037/1082-989X.8.4.448>
- Smith, E., Nichols, S., Lemke, J., Verdolini, K., Gray, S. D., Barkmeier, J., Dove, H. & Hoffman, H. (1994). Effects of voice disorders on patient lifestyle: Preliminary results. *NCVS Status and Progress Report*, 4, 237–248.
- Ubillos, S., Centeno, J., Ibañez, J. & Iraurgi, I. (2015). Protective and risk factors associated with voice strain among teachers in Castile and Leon, Spain: recommendations for voice training. *Journal of Voice*, 29(2), 261.e1–12. <https://doi.org/10.1016/j.jvoice.2014.08.005>
- van Houtte, E., Claeys, S., Wuyts, F. & van Lierde, K. (2011). The impact of voice disorders among teachers: vocal complaints, treatment-seeking behavior, knowledge of vocal care, and voice-related absenteeism. *Journal of Voice*, 25(5), 570–575. doi:10.1016/j.jvoice.2010.04.008
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer.

