

# Erweiterung der Global Scale of English (GSE) auf die Global Scale of Languages (GSL)

Teil 2: Anpassung der Lernziele  
für Deutsch an die GSL

November 2023

Ying Zheng, University of Southampton  
Catherine Doyle, Pearson  
David Booth, Pearson  
Mike Mayor, Pearson



---

# Inhaltsverzeichnis

<b>Kurzfassung.....</b>	<b>4</b>
<b>1. Einführung in die GSE- und GSL-Lernziele .....</b>	<b>5</b>
<b>2. Zweck der Studie .....</b>	<b>5</b>
<b>3. Methode.....</b>	<b>6</b>
3.1 Vergleichende Beurteilung und ihre Anwendungsfelder .....	6
3.2 Studiendesign.....	7
3.3 Lernziel – Übersetzungen: Vom Englischen ins Deutsche .....	8
3.4 Auswahl der Prüfer.....	9
3.5 Beschreibung des Datensatzes.....	10
<b>4. Ergebnisse .....</b>	<b>10</b>
4.1 Prüfer – Statistiken mit Informationsgewichtung .....	10
4.2 Lernziel – Statistiken mit Informationsgewichtung .....	12
4.2.1 Hören .....	14
4.2.2 Lesen .....	15
4.2.3 Sprechen .....	16
4.2.4 Schreiben .....	17
<b>5. Kombination spanischer und deutscher Datensätze.....</b>	<b>18</b>
<b>6. Diskussion und Schlussfolgerungen.....</b>	<b>20</b>
<b>Quellenangaben .....</b>	<b>21</b>
<b>Anhang: Demografische Daten der Prüfer.....</b>	<b>23</b>

---

## Kurzfassung

Die Global Scale of English (GSE) bietet ein detailliertes Instrument zur Beschreibung und Bewertung der Fortschritte und Leistungen von Englischlernenden. Pearson hat umfangreiche Forschungsarbeiten durchgeführt (siehe **Pearson**), um die GSE-Lernziele als Referenzskala für die Erweiterung der Can-do-Aussagen des Gemeinsamen Europäischen Referenzrahmens für Sprachen (Common European Framework of Reference for Languages, CEFR) aus dem Jahr 2001 zu nutzen und so den Bedürfnissen von mehr Lernenden gerecht zu werden.

Ziel dieser Studie war es, zu überprüfen, ob die Lernziele der neu eingeführten Global Scale of Languages (GSL) auf erwachsene Lernende von Deutsch als Fremdsprache (German-as-a-Foreign-Language, GFL) anwendbar sind. 320 GSE-Lernziele wurden ins Deutsche übersetzt. Ein Gremium aus 20 qualifizierten Prüfern aus einem Pool von GFL-Lehrkräften wurde aufgefordert, 25 vergleichende Beurteilungen (vB) pro Lernziel durchzuführen, was 8000 Datenpunkte ergab.

Eine Reihe von Analysen, einschließlich passender Statistiken für Prüfer und Testaufgaben, wurden durchgeführt. Es wurden starke Korrelationen zwischen den vB-Werten der Lernziele in der deutschen, spanischen und englischen Version sowie mit den ursprünglichen GSE-Werten festgestellt. Weitere Analysen der kombinierten spanischen und deutschen vB-Daten bestätigen die Übereinstimmung der GSE mit der GSL.

---

## 1. Einführung in die GSE- und GSL-Lernziele

Der GSE ist eine standardisierte Niveaustufe für Englisch, die von 10 bis 90 reicht und psychometrisch an den Gemeinsamen Europäischen Referenzrahmen für Sprachen (CEFR, Europarat, 2001) angelehnt ist. Es wurde eine Reihe von GSE-Lernzielen entwickelt, um die Fähigkeiten der Lernenden an jedem Punkt der Skala zu beschreiben, wobei das CEFR-Beschreibungsset einbezogen und erweitert wurde. Diese Lernziele wurden von Lehrkräften für Englisch als Fremdsprache (English as a Foreign Language, EFL) bewertet und anhand der Global Scale of English (de Jong, Mayor & Hayes, 2016) kalibriert. Im Gegensatz zum CEFR und einigen anderen Skalen, die das erreichte Niveau in groben Stufen beschreiben, zeigt die Global Scale of English auf, was ein Lernender an jedem Punkt der Skala in den Bereichen Sprechen, Hören, Lesen und Schreiben leisten kann, um eine detailliertere Beschreibung der zunehmenden Sprachkenntnisse zu liefern. Die Arbeit zur Entwicklung der GSE-Lernziele baut auf den von Brian North und dem Europarat bei der Erstellung des CEFR (North, 2000) durchgeföhrten Untersuchungen auf und erweitert diese. Die GSE-Lernziele wurden von Pearson English über mehrere Jahre hinweg in Zusammenarbeit mit über 6.000 Lehrkräften, ELT-Autoren und Sprachexperten aus der ganzen Welt entwickelt.

Der GSL für Französisch, Spanisch und Italienisch wurde im September 2023 mit dem Ziel eingeföhrt, die GSE-Erweiterung des CEFR für Lehrkräfte anderer Sprachen verfügbar und relevant zu machen. Zu Spanisch als Fremdsprache wurde eine Studie durchgeföhrt (Zheng, Doyle, Booth & Mayor, 2023), in der die Übereinstimmung zwischen den GSE-Werten und den spanischen Bewertungen derselben Lernziele validiert wurde, was zur Erstellung der Global Scale of Languages (GSL) führte.

## 2. Zweck der Studie

Um die GSL zu konsolidieren, werden in dieser Studie die Rangfolge der deutschen Übersetzungen der GSE-Lernziele verglichen, um zu erkennen, ob die bestehenden GSE-Werte auf erwachsene Lernende von Deutsch als Fremdsprache anwendbar sind, d. h., ob auf sie dieselbe Skala anwendbar ist.

---

Die Arbeitshypothese lautet: Da die GSE auf dem CEFR basiert, der seinerseits sprachneutral ist, wird angenommen, dass die Gesamtreihe folge sowohl mit der GSE als auch mit dem CEFR in hohem Maße korreliert. In diesem Projekt soll diese Hypothese mithilfe des Ansatzes der vergleichenden Beurteilung überprüft werden.

## 3. Methode

### 3.1 Vergleichende Beurteilung und ihre Anwendungsfelder

Bei der vergleichenden Beurteilung (vB) werden zwei Schülerarbeiten von einer Gruppe unabhängiger Prüfer holistisch beurteilt, um festzustellen, welche Arbeit das größere spezifizierte Gesamtkonstrukt aufweist. Das Ergebnis ist eine binäre Entscheidungsmatrix mit dem „Gewinner“ und dem „Verlierer“ jeder paarweisen Gegenüberstellung, die dann an das Bradley-Terry-Modell (Bradley & Terry, 1952) angepasst wird, um für jede Schülerarbeit Parameterwerte (Punktzahlen) und Standardfehler zu ermitteln. Der Parameterwert ermöglicht die Erstellung einer skalierten Rangfolge der Schülerarbeiten von „am besten“ bis „am schlechtesten“, die für Bewertungszwecke wie die Benotung verwendet werden kann.

Neben der Anwendung in britischen Prüfungsausschüssen, um auf die Vergleichbarkeit über verschiedene Prüfungsorganisationen hinweg (z. B. Fearnley, 2000; Gray, 2000), die Vergleichbarkeit von Standards im Zeitverlauf zu schauen und Standards aufrechtzuerhalten (z. B. Chambers & Cunningham, 2022), wurde die vB auch in einer Vielzahl von Bildungskontexten eingesetzt. Dazu gehören die Peer-Evaluation von Projektberichten von Studierenden im Bereich „Design Thinking“ (Mentzer, Lee, & Bartholomew, 2021), schriftliche Tests zum konzeptionellen Verständnis eines Mathematikkurses (Jones & Alcock, 2014), Lehrkraftbewertungen von summativen Statistik- und Englischbewertungen (Marshall, Shaw, Hunter, & Jones, 2020), Aufsätze (Steedle & Ferrara, 2016) und argumentative Texte (Lesterhuis, Verhavert, Coertjens, Donche, & De Maeyer, 2017). Pearson setzte die vB ein, um die Lernziele der Global Scale of English (GSE) für junge Lernende an das Leistungsniveau der Chinesischen Bewertungsskala für Englischkenntnisse (CSE) anzugeleichen, indem die Schwierigkeit der Deskriptoren in jedem Standard verglichen wurde (Pearson, 2020).

---

Die psychologische Grundlage für die vB besteht darin, dass Menschen ein Objekt mit einem anderen vergleichen können, aber Objekte isoliert nur unzuverlässig bewerten (Gill & Bramley, 2013; Thurstone, 1927). Bei den traditionellen analytischen Ansätzen bewerten die Lehrkräfte die Arbeiten der Studierenden individuell und absolut mithilfe von Rubriken, was zu unterschiedlichen Interpretationen und Anwendungen der Rubriken-Deskriptoren führen kann, ebenso wie zu der Möglichkeit, sich auf die Wahrnehmung der Arbeiten anderer Studierender zu stützen. Im Gegensatz dazu minimiert die vB den vergleichenden Einfluss detaillierter und spezifischer Rubriken (Pollitt, 2004), sie macht sich den vergleichenden Aspekt der Bewertung direkt zunutze und verzichtet auf Rubriken und eine Bewertung. Bisher wurde in der Literatur dargelegt, dass die vB hohe Anforderungen an die Validität, Zuverlässigkeit und Effizienz erfüllt.

### 3.2 Studiendesign

*NoMoreMarking* (Wheadon, 2019), ein vB-Tool, wurde zur Durchführung dieser Studie verwendet. Die Häufigkeit, mit der ein bestimmtes Objekt im Vergleich zu einem anderen beurteilt wird, ist ein wichtiges Element in einer vB-Studie. Verhavert, Vouwer, Donche und De Maeyer (2019) empfehlen 10 bis 30 Vergleiche pro Objekt, um eine akzeptable Zuverlässigkeit zu gewährleisten. In Übereinstimmung mit dieser Empfehlung wurden 25 Vergleiche pro Lernziel gesammelt, um ein verlässliches Studiendesign zu gewährleisten.

In dieser Studie haben wir 320 GSE-Lernziele für erwachsene Lernende ausgewählt. Die Stichprobengröße von 320 entspricht 30 % der Gesamtzahl der verfügbaren GSE-Lernziele. Bezuglich der Stichprobengröße und →auswahl sind im Allgemeinen 20 % die Mindestüberschneidung, die für die Angleichung der Skalen erforderlich ist (Kolen & Brennan, 2004). Die Stichprobe ist so geschichtet, dass sie sowohl für die Anzahl der Lernziele in jeder der vier Fertigkeiten als auch für die Anzahl in jedem CEFR-Niveau repräsentativ ist (siehe Tabelle 1 unten).

Tabelle 1: Verteilung der Lernziele

CEFR/GSE	Hören	Lesen	Sprechen	Schreiben	GESAMT	% der Datenbank
Unter A1 (10–21)	3	3	10	4	20	34 %
A1 (22–29)	5	5	14	8	32	27 %
A2 (30–35)	6	6	17	10	39	30 %
A2+ (36–42)	6	6	16	10	38	27 %
B1 (43–50)	7	7	18	11	43	35 %
B1+ (51–58)	7	7	18	11	43	33 %
B2 (59–66)	7	7	19	11	44	28 %
B2+ (67–75)	5	5	14	8	32	27 %
C1 (76–84)	3	3	10	4	20	28 %
C2 (85–90)	1	1	4	3	9	47 %
<b>INSGESAMT</b>	<b>50</b>	<b>50</b>	<b>140</b>	<b>80</b>	<b>320</b>	<b>30 %</b>
<b>% der Datenbank</b>	<b>26 %</b>	<b>35 %</b>	<b>28 %</b>	<b>33 %</b>	<b>30 %</b>	

### 3.3 Lernziel – Übersetzungen: Vom Englischen ins Deutsche

Die 320 GSE-Lernziele wurden von einem Übersetzungsbüro mit Erfahrung in der Übersetzung von Bildungsmaterialien ins Deutsche übersetzt. Der Agentur wurden die offiziellen CEFR-Übersetzungen aus dem Englischen ins Deutsche des Europarats sowie die Stilrichtlinien von Pearson als Leitfaden zur Verfügung gestellt. Der Übersetzungsprozess durchlief mehrere Phasen:

- Glossar der wichtigsten Begriffe: Übersetzung durch die Agentur,
- Glossar der wichtigsten Begriffe: Überprüfung durch die deutschsprachigen Mitarbeiter von Pearson,
- Erste Übersetzungsrounde durch die Agentur,
- Überprüfung und Korrekturen durch einen zweiten Übersetzer innerhalb der Agentur,
- Überprüfung durch die deutschsprachigen Mitarbeiter von Pearson,
- Überprüfung durch einen externen deutschen Redakteur, der von Pearson beauftragt wurde.

---

Um eine Verknüpfung mit der spanischen Alignment-Studie (Zheng et al., 2023) herzustellen, wurden in dieser Studie dieselben Lernziele verwendet, mit Ausnahme von 6 (2 Sprechen, 4 Schreiben), die durch andere Lernziele auf demselben CEFR-Niveau ersetzt wurden.

Dies war das Ergebnis einer qualitativen Überprüfung, bei der festgestellt wurde, dass diese 6 Wörter eine grammatisch Nuance aufweisen, wodurch sie für die Verwendung in anderen Sprachen als Englisch ungeeignet sein könnten.

Zum Beispiel:

*Kann sehr einfache Konnektoren wie „und“, „aber“, „so“ und „dann“ verwenden.*

*Kann sehr einfache Ereignisse in der Vergangenheit mit einfachen Verbindungswörtern (z. B. „dann“, „als nächstes“) beschreiben.*

### 3.4 Auswahl der Prüfer

Die Prüfer waren allesamt erfahrene GFL-Lehrkräfte. Sie wurden aus den folgenden drei Bereichen rekrutiert:

- Leitende Prüfer für die Pearson Edexcel GCSE- und/oder A-Level-Qualifikation Deutsch (Sekundarschul-/College-Abschlüsse im Vereinigten Königreich),
- Deutschlehrkräfte im Vereinigten Königreich,
- Deutschlehrkräfte in Polen.

174 Personen bekundeten ihr Interesse an der Teilnahme an der Studie und gaben einige Hintergrundinformationen an. Auf der Grundlage ihrer Erfahrung im Unterricht von erwachsenen Lernenden sowie ihrer Vertrautheit mit dem CEFR wurden 20 Prüfer für das Projekt ausgewählt. Zudem wurde darauf geachtet, eine Gruppe von Prüfern zusammenzustellen, die im Hinblick auf Geschlecht, Nationalität und Erfahrung so vielfältig wie möglich ist (siehe Anhang für die demografischen Daten der Prüfer). Die Prüfer erhielten eine schriftliche Anleitung zur Aufgabe und zur Plattform in englischer und/oder polnischer Sprache. Anschließend wurden sie gebeten, eine vergleichende Beurteilung auf der Grundlage dieser Frage vorzunehmen: „Welches dieser Lernziele beschreibt eine schwierigere Fähigkeit für einen Sprachenlernenden?“

### 3.5 Beschreibung des Datensatzes

Tabelle 2: Anzahl der Lernziele und Vergleiche für jede Fähigkeit

Fähigkeit	Lernziele für Deutsch	Gesamtzahl der Beurteilungen
Hören	50	1.250
Lesen	50	1.250
Sprechen	140	3.500
Schreiben	80	2.000
<b>GESAMT</b>	<b>320</b>	<b>8.000</b>

## 4. Ergebnisse

Bei der vergleichenden Beurteilung wird die Zuverlässigkeit der Trennung (Scale Separation Reliability, SSR) als Indikator für die Zuverlässigkeit verwendet, in diesem Fall für die Zuverlässigkeit der Rangfolge der Lernziele, die sich aus der vB-Aktivität ergibt.

Die SSR wird auf einer Skala von 0 bis 1 angegeben, wobei Werte über 0,90 auf eine sehr zuverlässige vB-Skala hinweisen. Tabelle 3 zeigt die SSR für alle vier Fertigkeiten, wobei Schreiben die niedrigste Zuverlässigkeit (0,93) und Lesen die höchste Zuverlässigkeit (0,95) aufweist.

Tabelle 3: Zuverlässigkeit der Trennung

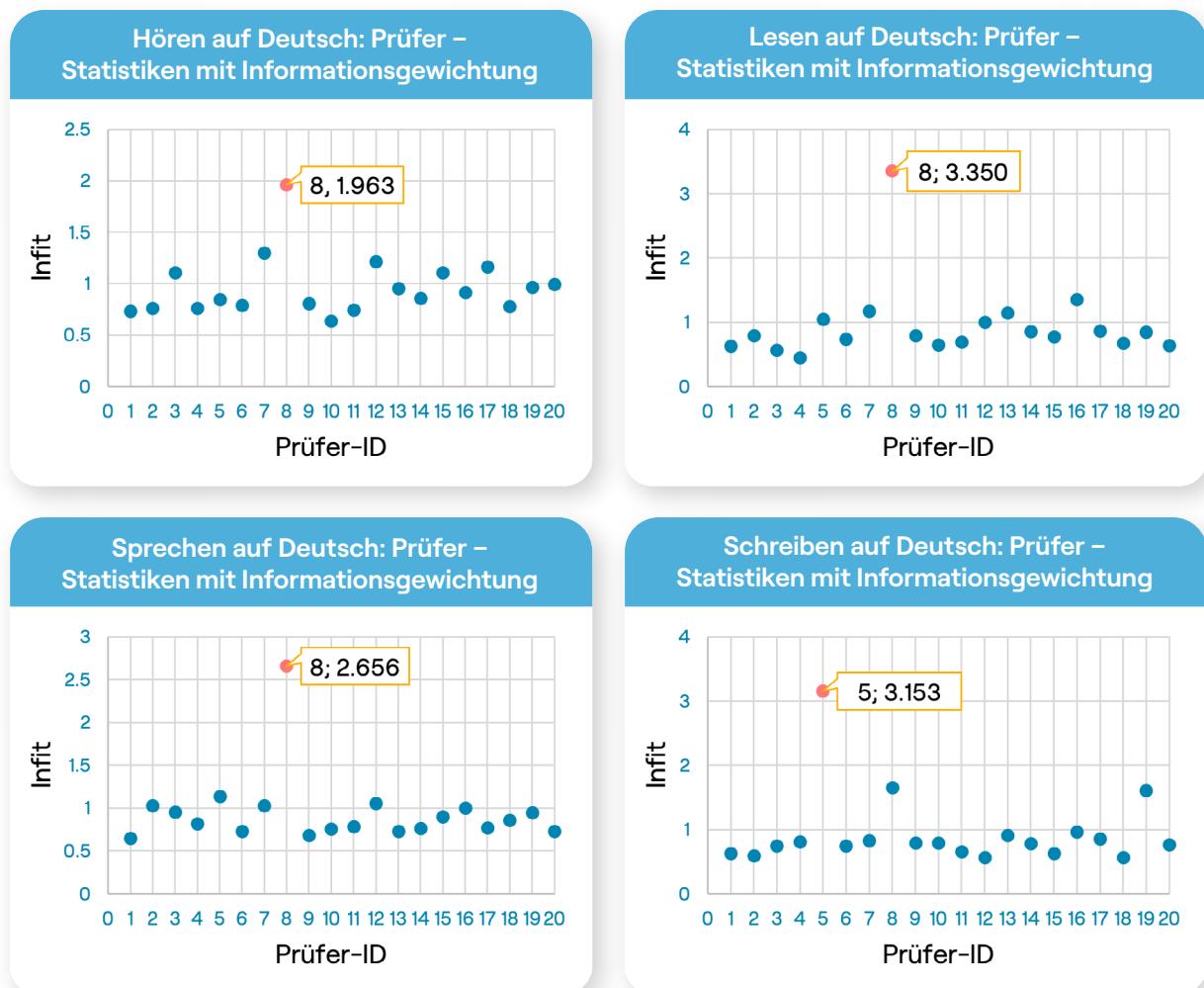
Hören	Lesen	Sprechen	Schreiben
0,94	0,95	0,94	0,93

### 4.1 Prüfer – Statistiken mit Informationsgewichtung

Passende Statistiken wurden sowohl für die Prüfer als auch für die Items (d. h. die Lernziele) berechnet, die in dieser vB-Übung verwendet wurden. Prüfer mit einer Informationsgewichtung (Infit), die mehr als zwei Standardabweichungen über der durchschnittlichen Informationsgewichtung liegt, wurden ausgeschlossen, da dies darauf hindeutet, dass sie möglicherweise widersprüchliche Beurteilungen abgeben oder nicht mit dem Konsens der anderen Prüfer übereinstimmen.

Wie die vier folgenden Abbildungen zeigen, wies Prüfer Nr. 8 hohe Statistiken mit Informationsgewichtung für drei Fertigkeiten auf: Hören, Lesen und Sprechen. Prüfer Nr. 5 wies hohe Statistiken mit Informationsgewichtung für das Schreiben auf. Die entsprechenden Verhältnisdaten wurden daher aus der weiteren Analyse und Berichterstattung ausgeklammert.

*Abbildung 1 – 4: Prüfer – Statistiken mit Informationsgewichtung (vier Fertigkeiten)*



---

## 4.2 Lernziel – Statistiken mit Informationsgewichtung

In den folgenden Abschnitten werden die Statistiken mit Informationsgewichtung für die vier Fertigkeiten aufgeführt, die im Rahmen des Lernziels erworben wurden. Die Abbildungen 5–8 zeigen die Streudiagramme für jede Fähigkeit, wobei die Y-Achse die Statistiken mit Informationsgewichtung und die X-Achse die Anzahl der Items angibt. Lernziele mit Statistiken mit Informationsgewichtung außerhalb des zufriedenstellenden Bereichs sind in den entsprechenden Abbildungen rot hervorgehoben. Wie zu erkennen ist, sind zwei Lernziele im Bereich Hören, eines im Bereich Lesen, drei im Bereich Sprechen und fünf im Bereich Schreiben hervorgehoben.

Diese Lernziele wurden dann einer qualitativen Überprüfung unterzogen. Wie zu erwarten war, bestand bei der Betrachtung der Varianz der Infit-Werte, wie in den Abbildungen 5–8 dargestellt, kein allgemeines Merkmal, das die Infit-Werte erklären würde.

In einigen Fällen könnte dies daran liegen, dass die Aufgabe auf Deutsch schwieriger ist. Ein Beispiel für das folgende Lernziel „Hören“:

*Kann die Kardinalzahlen von 1 bis 20 verstehen.*

*Can understand cardinal numbers from 1 to 20.*

*GSE 10*

Im Deutschen folgt das Muster für Kardinalzahlen nach 10 nicht dem gleichen Muster wie im Englischen. Obwohl die Diskussionsteilnehmer nicht gebeten wurden, einen Vergleich mit dem Englischen anzustellen, könnte dieser Aspekt einen Einfluss auf die Bewertung dieses Lernziels gehabt haben.

Auch bei den Lernzielen für das Schreiben konnte kein allgemeines Muster festgestellt werden, obwohl zwei von ihnen Konzepte wie Brüche und zusammengesetzte Buchstaben einführten, die für die Diskussionsteilnehmer verwirrend gewesen sein könnten. Siehe die nachstehenden Beispiele:

*Kann konsequent mit zusammenhängenden Buchstaben schreiben.*

*Can write consistently with joined-up letters.*

*GSE 20*

*Kann Brüche sowohl mit Ziffern als auch mit Wörtern schreiben.*

*Can write fractions using both digits and words.*

*GSE 50*

---

Da nur eine kleine Anzahl von Lernzielen Abweichungen aufwies und es dafür keine allgemeine Erklärung gab, wurde beschlossen, alle Lernziele der Studie in der Lernziel-Datenbank zu belassen.

Die Tabellen 4–7 zeigen die Korrelationen zwischen den vB-Werten, die aus den englischen, spanischen und deutschen Versionen der Lernziele generiert wurden, sowie mit den ursprünglichen Werten der Global Scale of English. Die Ergebnisse sind zufriedenstellend, wie die hohen Korrelationen zwischen den Ergebnissen zeigen.

Die Lernziele, die sich nach der Übersetzung als mehrdeutig erwiesen, wurden durch Lernziele mit ähnlichen GSE-Werten ersetzt. Die Ersatzdaten, die keine Entsprechung in der spanischen/englischen Version haben, werden in der entsprechenden Korrelationsanalyse nicht berücksichtigt.

## 4.2.1 Hören

Abbildung 5: Hören – Lernziel – Statistiken mit Informationsgewichtung

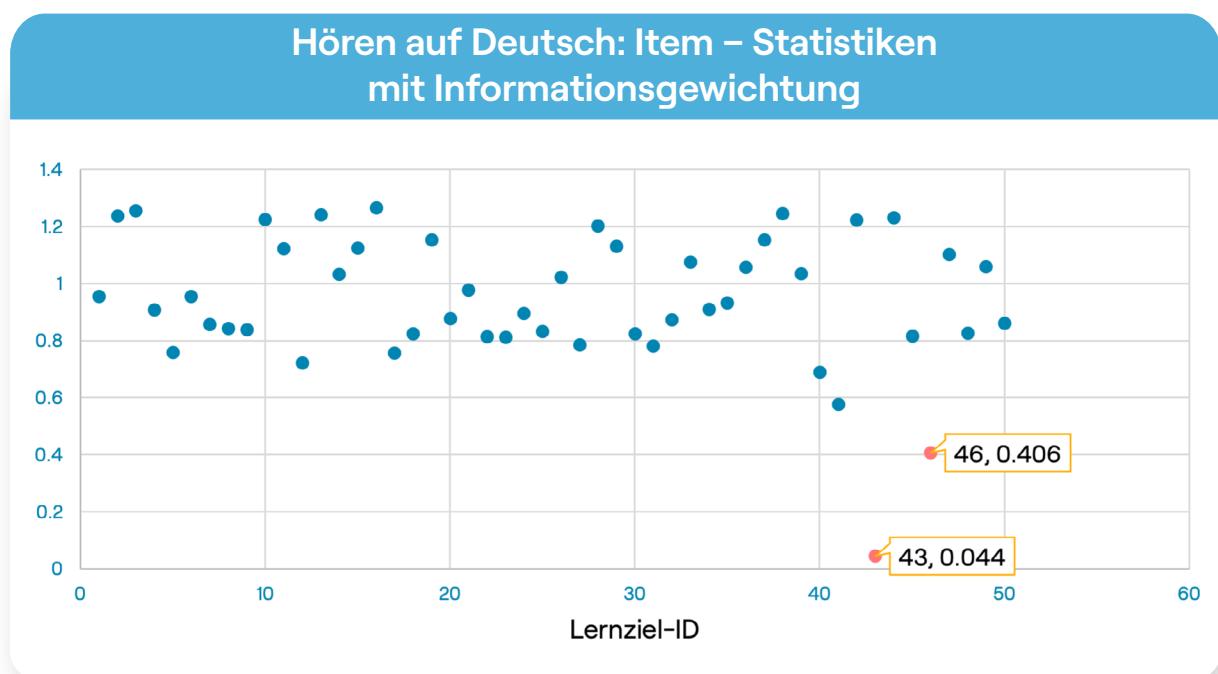


Tabelle 3: Hören – Vergleich der vB-Schätzungen

	GSE	Deutsche vB Skalierte Punktzahl	Spanische vB Skalierte Punktzahl	Englische vB Skalierte Punktzahl
GSE	1			
Deutsche vB Skalierte Punktzahl	0,919	1		
Spanische vB Skalierte Punktzahl	0,955	0,895	1	
Englische vB Skalierte Punktzahl	0,931	0,858	0,928	1

## 4.2.2 Lesen

Abbildung 6: Lesen – Lernziel – Statistiken mit Informationsgewichtung

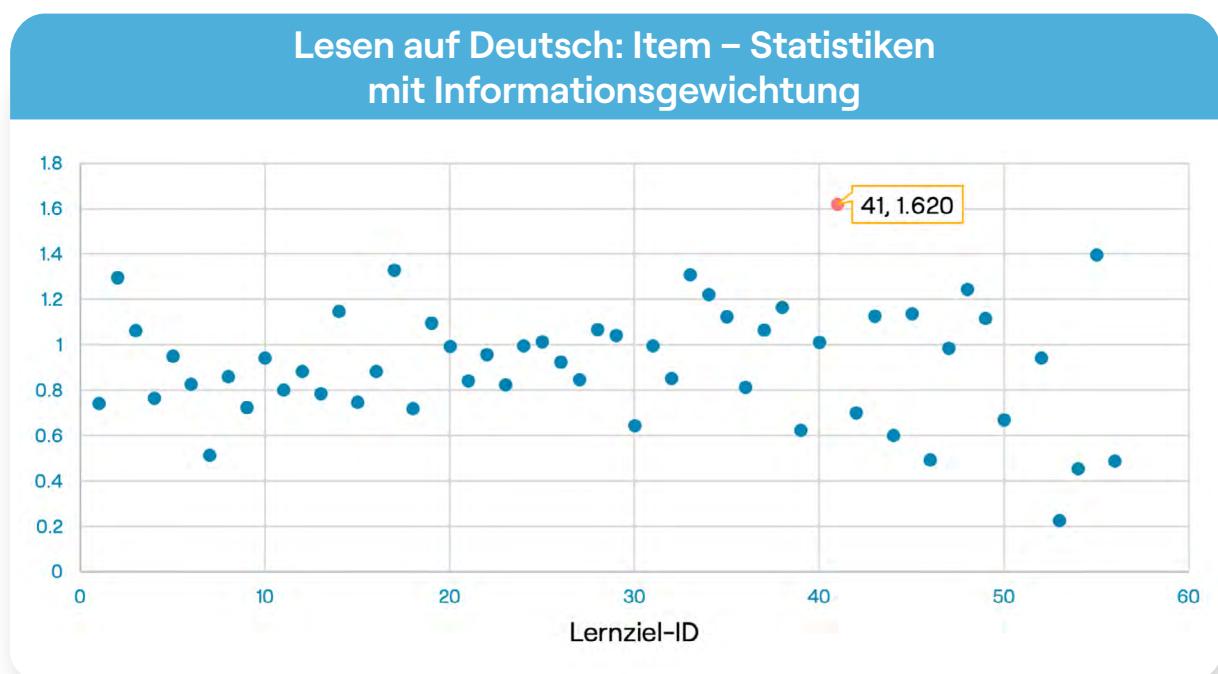


Tabelle 5: Lesen – Vergleich der vB-Schätzungen

	GSE	Deutsche vB Skalierte Punktzahl	Spanische vB Skalierte Punktzahl	Englische vB Skalierte Punktzahl
GSE	1			
Deutsche vB Skalierte Punktzahl	0,935	1		
Spanische vB Skalierte Punktzahl	0,944	0,928	1	
Englische vB Skalierte Punktzahl	0,912	0,906	0,920	1

### 4.2.3 Sprechen

Abbildung 7: Sprechen – Lernziel – Statistiken mit Informationsgewichtung

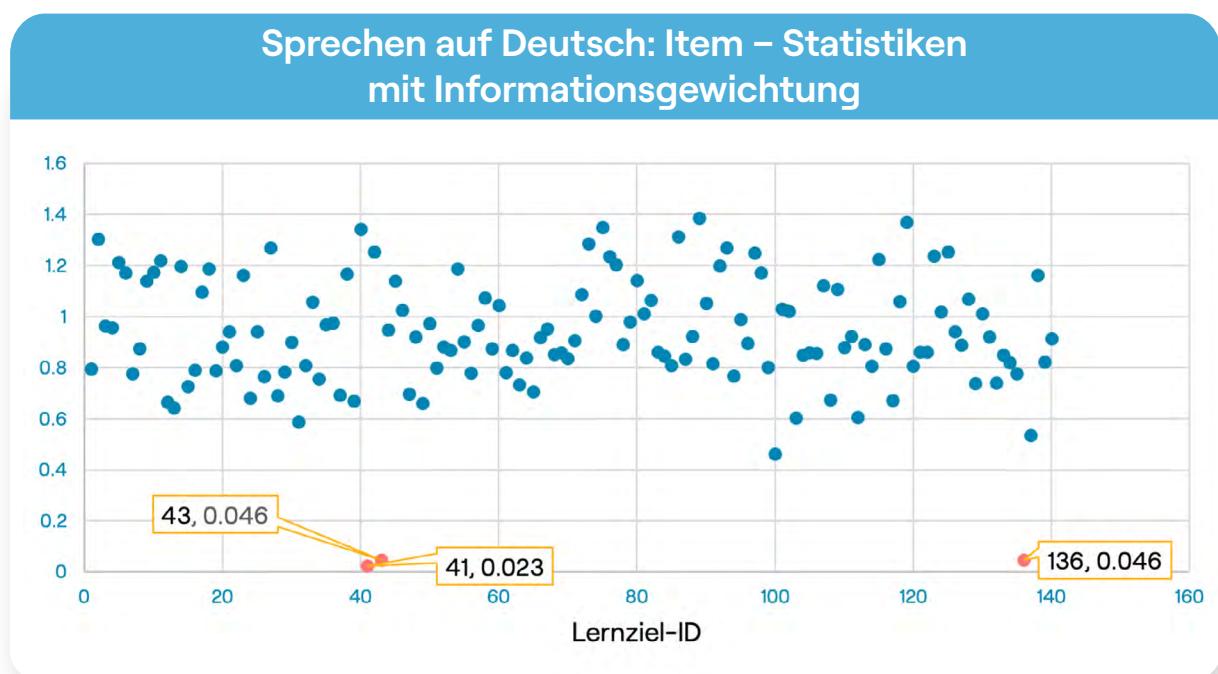


Tabelle 6: Sprechen – Vergleich der vB-Schätzungen

	GSE	Deutsche vB Skalierte Punktzahl	Spanische vB Skalierte Punktzahl	Englische vB Skalierte Punktzahl
GSE	1			
Deutsche vB Skalierte Punktzahl	0,902	1		
Spanische vB Skalierte Punktzahl	0,915	0,916	1	
Englische vB Skalierte Punktzahl	0,911	0,866	0,867	1

#### 4.2.4 Schreiben

Abbildung 8: Schreiben – Lernziel – Statistiken mit Informationsgewichtung

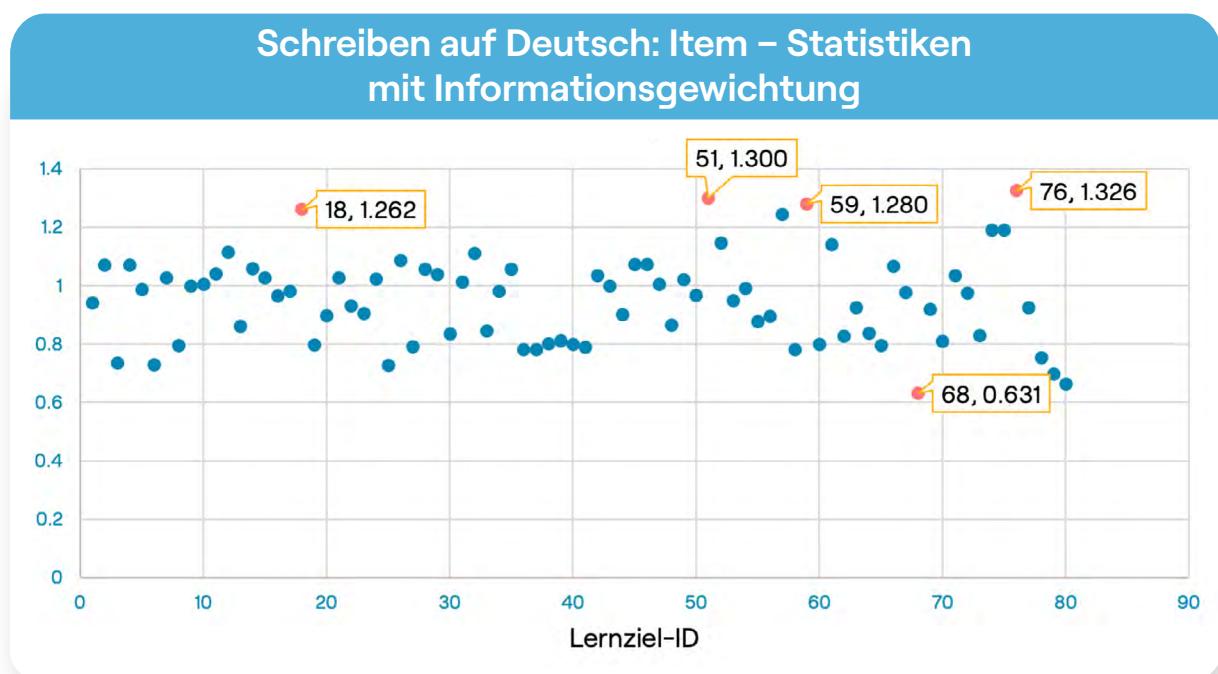


Tabelle 7: Schreiben – Vergleich der vB-Schätzungen

	GSE	Deutsche vB Skalierte Punktzahl	Spanische vB Skalierte Punktzahl	Englische vB Skalierte Punktzahl
GSE	1			
Deutsche vB Skalierte Punktzahl	0,928	1		
Spanische vB Skalierte Punktzahl	0,948	0,912	1	
Englische vB Skalierte Punktzahl	0,889	0,866	0,888	1

## 5. Kombination spanischer und deutscher Datensätze

Um die Global Scale of Languages zu konsolidieren, wurden auf der Grundlage der kombinierten spanischen und deutschen Datensätze Transformationsgleichungen von vB-Skalenwerten zu GSL für jede Sprachfertigkeit erstellt. Die Abbildungen 9 bis 12 zeigen die Transformationsgleichungen der kombinierten spanischen und deutschen vB-Werte in die GSE/GSL.

Im Vergleich zu den Transformationsgleichungen, die nur auf der Grundlage der spanischen Daten erstellt wurden, ist festzustellen, dass die Varianzen in der GSE, die durch die vB-Werte erklärt werden können, in allen vier Fertigkeiten zugenommen haben. Genauer gesagt,

- Hören, R-Quadrat: 0,912 (Daten für Spanisch); R-Quadrat 0,921 (Daten für Spanisch und Deutsch zusammen),
- Lesen, R-Quadrat: 0,831 (Daten für Spanisch); R-Quadrat 0,915 (Daten für Spanisch und Deutsch zusammen),
- Sprechen: R-Quadrat: 0,831 (Daten für Spanisch); R-Quadrat 0,880 (Daten für Spanisch und Deutsch zusammen),
- Schreiben: R-Quadrat: 0,791 (Daten für Spanisch); R-Quadrat 0,887 (Daten für Spanisch und Deutsch zusammen).

Abbildung 9: Hören

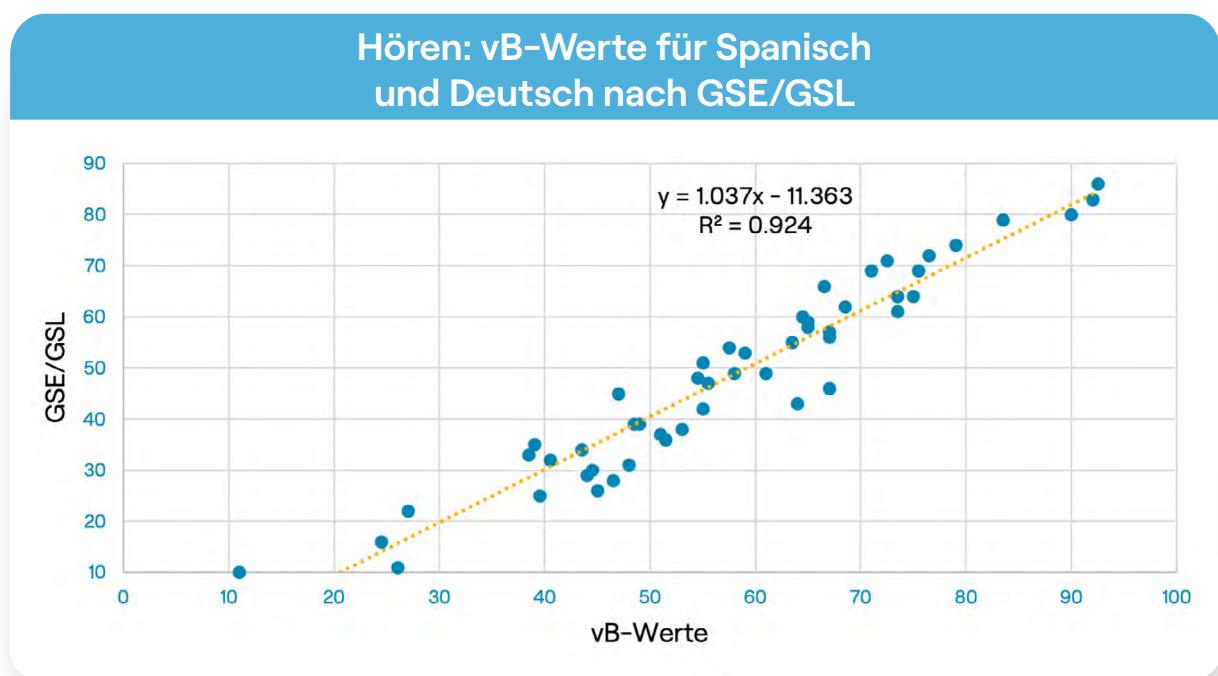


Abbildung 10: Lesen

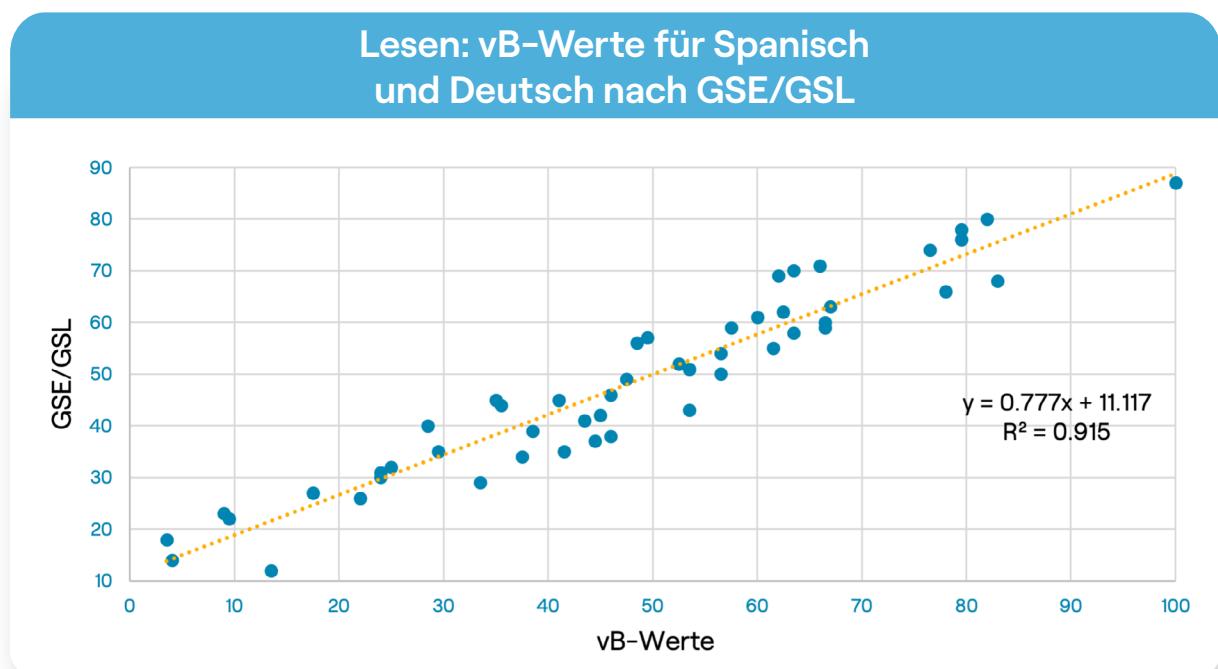


Abbildung 11: Sprechen

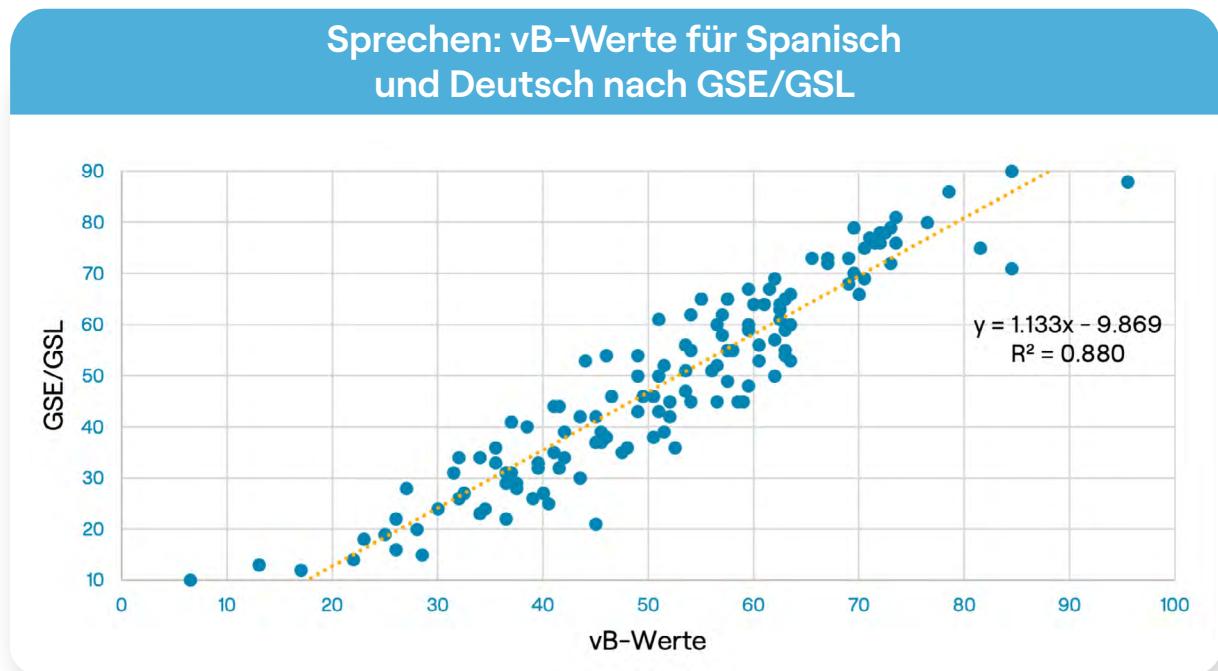
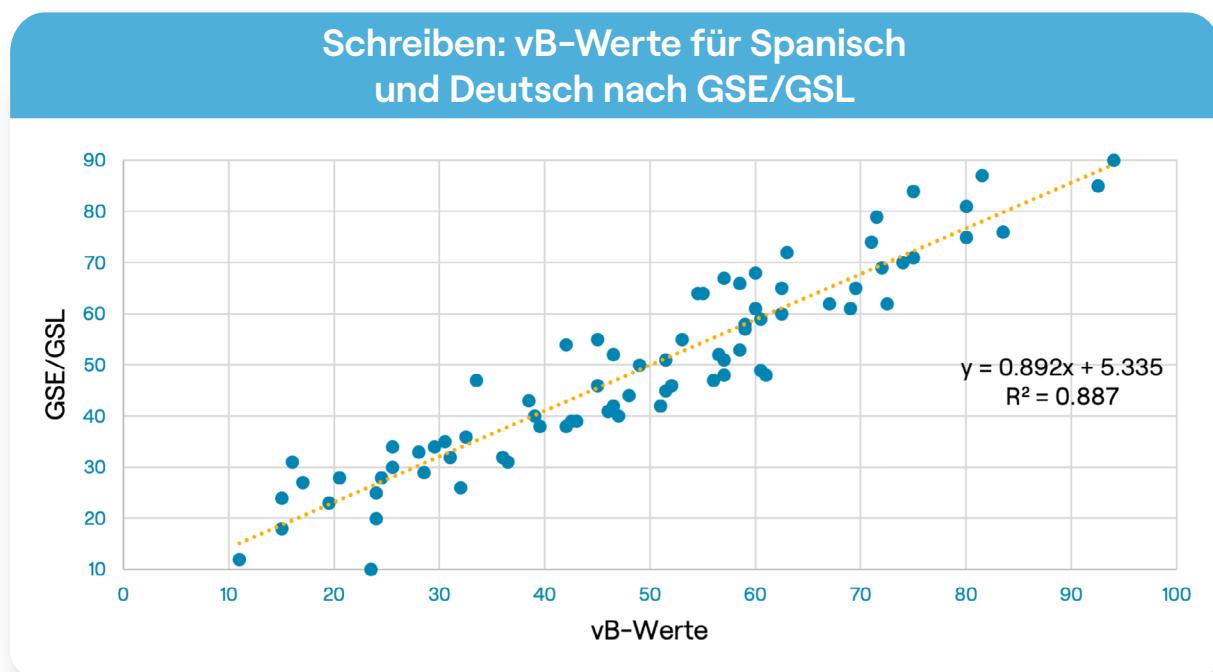


Abbildung 12: Schreiben



## 6. Diskussion und Schlussfolgerungen

Die Ergebnisse dieser vB-Studie zeigen hohe Korrelationen zwischen den Niveaustufen der gleichen Lernziele in Englisch, Spanisch und Deutsch. Der CEFR selbst ist ein sprachneutraler Rahmen, der „für verschiedene Kontexte angepasst und für alle Sprachen verwendet werden kann“ (Europarat), und seit seiner Entwicklung im Jahr 2001 wurde er in 40 Sprachen übersetzt (ebd.). Die Arbeit von Pearson zur Erweiterung des CEFR und zur Schaffung der GSE wurde ursprünglich im Zusammenhang mit Englisch als Fremdsprache (EFL) konzipiert. Es wurde jedoch davon ausgegangen, dass diese Erweiterung auch für Lehrkräfte und Lernende anderer Sprachen relevant und nützlich sein könnte.

Die in diesem Bericht beschriebene vB-Studie liefert Belege dafür, dass die kommunikativen, funktionalen Sprachhandlungen, die in Can-do-Aussagen ausgedrückt werden, sowohl im Englischen als auch im Deutschen gleichwertig sind, d. h., dass beide auf der gleichen Skala eingeordnet werden können. Die Analyse dieses deutschen Datensatzes ergänzt die Arbeit, die im Rahmen der früheren GSL-Studie umgesetzt wurde, und liefert weitere Belege für die Validität der GSL. Es wird weiter daran gearbeitet, weitere Sprachen in die Validierung der GSL einzubeziehen.

---

## Quellenangaben

- Bradley, R. A. and Terry, M. E. (1952). *Rank analysis of incomplete block designs. I. The method of paired comparisons*. Biometrika 39 324–345.
- Chambers, L., & Cunningham, E. (2022). *Exploring the Validity of Comparative Judgement: Do Judges Attend to Construct-Irrelevant Features?* Frontiers in Education (7).
- Council of Europe (2001). *Common European framework of reference for languages: learning, teaching, assessment*. Cambridge: Cambridge University Press.
- de Jong, J., Mayor, M., & Hayes, C. (2016). *Developing Global Scale of English Learning Objectives aligned to the Common European Framework*. Available at: <https://www.pearson.com/languages/why-pearson/the-global-scale-of-english/resources.html>
- Fearnley, A. (2000). A comparability study in GCSE mathematics. A study based on the summer 1998 examination. In *Assessment and Qualifications Alliance* (Northern Examinations and Assessment Board). Manchester: Joint Forum for the GCSE and GCE.
- Gill, T., & Bramley, T. (2013). How accurate are examiners' holistic judgements of script quality?. *Assessment in Education: Principles, Policy & Practice*, 20(3), 308–324.
- Gray, E. (2000). *A comparability study in GCSE science 1998. A study based on the 1998 summer examination*. Organised by Oxford, Cambridge and RSA Examinations (Midland Examining Group) on behalf of the joint forum for GCSE and GCE.
- Jones, I., & Alcock, L. (2014). Peer assessment without assessment criteria. *Studies in Higher Education*, 39(10), 1774–1787.
- Kolen, M. J., & Brennan R. L. (2004). *Test equating, scaling, and linking: Methods and practices*. 2nd. New York: Springer.
- Lesterhuis, M., Verhavert, S., Coertjens, L., Donche, V., & De Maeyer, S. (2017). Comparative judgement as a promising alternative to score competences. In *Innovative practices for higher education assessment and measurement* (pp. 119–138). IGI Global.

- 
- Marshall, N., Shaw, K., Hunter, J., & Jones, I. (2020). Assessment by comparative judgement: An application to secondary statistics and English in New Zealand. *New Zealand Journal of Educational Studies*, 55, 49–71.
- Mentzer, N., Lee, W., & Bartholomew, S. R. (2021). Examining the Validity of Adaptive Comparative Judgment for Peer Evaluation in a Design Thinking Course. In *Frontiers in Education* (p. 492). Frontiers.
- North, B. (2000). *The development of a common framework scale of language proficiency*. New York: Peter Lang.
- Pollitt, A. (2004). *Let's stop marking exams*, International Association for Educational Assessment Conference. Philadelphia PA.
- Steedle, J. T., & Ferrara, S. (2016). Evaluating comparative judgment as an approach to essay scoring. *Applied Measurement in Education*, 29(3), 211–223.
- Pearson technical report (2020): Aligning Global Scale of English-Young Learner to the CSE. Available at <https://m.i21st.cn/elt/15934.html>
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological review*, 34(4), 273.
- Verhavert, S., Bouwer, R., Donche, V., & Maeyer, S. D. (2019). A meta-analysis on the reliability of comparative judgement. *Assessment in Education: Principles, Policy & Practice*, 26(5), 541–562.
- Wheadon, C. (2019). *No More Marking* [Computer Software]. Retrieved from <https://www.nomoremarking.com/>
- Zheng, Y., Doyle, C., Booth, D., & Mayor, M. (2023). Extending the Global Scale of English (GSE) to the Global Scale of Languages (GSL): Aligning Spanish Learning Objectives to the GSL. <https://www.pearson.com/content/dam/one-dot-com/one-dot-com/pearson-languages/en-gb/pdfs/research-report-extending-the-global-scale-of-english-gse-to-the-global-scale-of-languages-gsl.pdf>

---

## Anhang: Demografische Daten der Prüfer

Staatsangehörigkeit	Anzahl
Polnisch	15
Britisch	2
Britisch und Deutsch	3
<b>Gesamtzahl</b>	<b>20</b>

Geschlecht	Anzahl
Weiblich	15
Männlich	5
<b>Gesamtzahl</b>	<b>20</b>

Jahre Erfahrung im Deutschunterricht	Anzahl
2–5 Jahre	4
5–10 Jahre	1
> 10 Jahre	15
<b>Gesamtzahl</b>	<b>20</b>

Vertrautheit mit CEFR	Anzahl
Vertiefte Kenntnisse	16
Allgemeines Verständnis	3
Ist bekannt	1
<b>Gesamtzahl</b>	<b>20</b>

<b>Andere unterrichtete Sprachen</b>	<b>Anzahl*</b>
Französisch	<b>4</b>
Englisch	<b>6</b>
Niederländisch	<b>2</b>
Polnisch	<b>2</b>
Latein	<b>1</b>

\* 11 Teilnehmer hatten mindestens eine Sprache unterrichtet

<b>Altersgruppe(n) im (Deutsch-) Unterricht</b>	<b>Anzahl</b>
Erwachsene (über 18)	<b>15</b>
Sekundarstufe II / College / 6. Klasse (15–19)	<b>19</b>
Sekundarstufe I (12–15)	<b>9</b>
Obere Grundschule (9–12)	<b>3</b>
Untere Grundschule (6–9)	<b>3</b>
Vorschule (3–5)	<b>0</b>







