

Forschungsbericht Zwanzigeins e.V., Stand: 16. Juli 2025

Peter Morfeld

Zwanzigeins e.V. (<https://zwanzigeins.jetzt/>)

Bochum/Schwerte

**Sekundäranalyse von frühen pädagogischen Studien zur  
Auswirkung der traditionell-verdrehten und einer  
stellenwertgerechten Zahlensprechweise im Deutschen auf  
arithmetische Fähigkeiten von Schulkindern (Schellenberger 1953,  
Zehner 1955, Ammareller 2006)**

Korrespondenzadresse

PD Dr. Peter Morfeld

Schlossweide 7

58239 Schwerte

[peter.morfeld@zwanzigeins.jetzt](mailto:peter.morfeld@zwanzigeins.jetzt)

## **Zusammenfassung**

Zu drei empirischen pädagogischen Studien (Schellenberger 1953, Zehner 1955, Ammareller 2006) wurde eine sekundäre statistische Analyse mit Tabellenauswertung und logistischer Regression durchgeführt, in der die Fehlerhäufigkeit unter Schülern bei stellenwertgerechter und bei traditionell-verdrehter Sprechweise der Zahlen verglichen wurde. Diese drei frühen didaktischen Arbeiten beschreiben einheitlich eine Reduktion der Fehlerzahl um 40% bis 50% bei Einsatz einer stellenwertgerechten Sprechweise. Dies Ergebnis ist insbesondere wichtig, da zwei Arbeiten (Schellenberger 1953, Ammareller 2006) auch Kopfrechenaufgaben umfassten, also nicht nur Zahlendiktate, und zwei Arbeiten (Zehner 1955, Ammareller 2006) die Wirkung einer längeren Einübungszeit in die stellenwertgerechte Sprechweise beurteilen konnten. Die vorliegende Sekundäranalyse der drei frühen didaktischen Studien erhärtet somit die plausible Hypothese, dass die Verwendung einer unverdrehten Zahlensprechweise im Mathematikunterricht zu deutlichen Vorteilen führt.

## **Schlagwörter**

Didaktische Studien, Zahlensprechweise im Mathematikunterricht, Fehlerhäufigkeitsanalyse, Zahlendiktat, Kopfrechnen

## **Gliederung**

1. Einleitung

2. Material und Methodik

3. Schellenberger 1953

3.1 Versuchsbeschreibung

3.2 Ergebnisse

3.3 Kommentar

4. Zehner 1955

4.1 Versuchsbeschreibung

4.2 Ergebnisse

4.3 Kommentar

5. Ammareller 2006

5.1 Versuchsbeschreibung

5.2 Ergebnisse

5.3 Kommentar

6. Überblick und Diskussion

7. Literaturverzeichnis

## 1. Einleitung

Im Deutschen lesen und schreiben wir von links nach rechts. Das stimmt aber nicht für die Zahlen. Anders als z.B. in den romanischen Sprachen, im Englischen, im Ukrainischen, im Russischen, im Türkischen oder im Chinesischen, spricht man im Deutschen die Zahl 123 nicht von links nach rechts als „hundert-zwanzig-drei“, sondern in verdrehter Abfolge als „hundert-drei-und-zwanzig“. Das widerspricht der Logik des Stellenwertsystems, d.h. der geordneten Zahlenfolge von Hunderter-Zehner-Einer.

Die gemeinnützige Initiative „Zwanzigeins e.V.“ möchte eine unverdrehte, d.h. eine stellenwertgerechte Zahlensprechweise im Deutschen populär und gesellschaftsfähig machen (Zwanzigeins e.V. 2025a). Sie soll als eine richtige Sprechweise allgemein anerkannt werden. Die von Zwanzigeins e.V. empfohlene Form der unverdrehten Sprechweise („zehneins“, Zwanzigeins e.V. 2025b) kann mit Hilfe der Zahlen 11, 14, 21, 30, 46 wie folgt erläutert werden: zehneins, zehnvier, zwanzigeins, dreißig, vierzigsechs. Neben der traditionell-verdrehten Sprechweise ist diese kollisionsfrei verwendbar.

Die Zwanzigeins-Initiative stellt als erste von drei Forderungen auf (Forderung A auf der Startseite von Zwanzigeins e.V. 2025a):

„Die für den Schulunterricht verantwortlichen gesellschaftlichen Gruppen, Gremien, Ministerien und Behörden sind angehalten, *empirische Untersuchungen* in Auftrag zu geben, in denen erforscht werden soll, inwieweit die deutsche Zahlensprechweise im Mathematikunterricht an Grundschulen für viele Schüler ein Lernhindernis darstellt.“

Daher wurden und werden aktuell zu diesem Thema pädagogische Arbeiten mit Hilfe der Zwanzigeins-App durchgeführt (Zwanzigeins 2025c). Allerdings gab es auch bereits früher Bemühungen im didaktischen Bereich, die aber mit unzureichenden Analysen der erhobenen Daten endeten (Zwanzigeins 2025c). Diese Arbeiten sind auch schlecht erreichbar<sup>1</sup>, und es gibt keine systematische und gemeinsame Darstellung. In diesem Bericht werden diese Arbeiten ausführlich vorgestellt und kommentiert. Zudem werden als sekundären Auswertung zu diesen Studien zufallskritische Analysen nach einheitlicher Methodik vorgenommen. Ziel ist es, diese

---

<sup>1</sup> Prof. Dr. Lothar Gerritzen ermöglichte den Zugang zu den drei Originalarbeiten, die nun vollständig zur Verfügung stehen.

frühen empirischen Bemühungen so aufzubereiten, dass sie in eine Diskussion um die Vorteile einer stellenwertgerechten Zahlensprechweise eingebracht werden können.

## 2. Material und Methodik

Drei empirische pädagogische Studien, in denen die Fehlerhäufigkeiten unter Schülern bei stellenwertgerechter und bei traditionell-verdrehter Sprechweise der Zahlen verglichen wurden und deren Basisdaten zumindest bis in eine gewisse Tiefe verfügbar oder rekonstruierbar waren, habe ich in die sekundäre statistische Analyse einbezogen: Schellenberger 1953, Zehner 1955 und Ammareller 2006. Darüber hinaus konnte ich keine weiteren Arbeiten zu dem Thema identifizieren, deren Basisdaten für eine zufallskritische Analyse zumindest in grober Rasterung verfügbar sind oder verfügbar gemacht werden könnten.

Einzeldaten zu Schulkindern waren zu keiner der drei verwendeten Arbeiten erreichbar, sondern allein aggregierte Daten. Und diese oft auch nicht als Absolut- sondern nur in Form von Prozentwerten. Ich rekonstruierte aus den vorliegenden Angaben die zugehörigen Zähler und Nenner und stellte diese Daten in Form von 2x2-Tabellen (4-Felder-Tabellen, siehe: Hartung et al. 2009, S. 411-419) zusammen, die den beobachteten Zusammenhang von Sprechweise und Fehlerhäufigkeit abbilden. Kovariablen zum Aufteilen der Tabellen in Untertabellen standen nicht zur Verfügung.

In den Ergebniskapiteln werden die aufgebauten 2x2-Tabellen wie folgt dargestellt, erläutert am Beispiel „3.2.1.1 Addition, leicht“ aus der Sekundäranalyse zu Schellenberger 1953:

	unverdr.	verdr.	Total
<b>Fehler</b>	4	7	11
<b>in %</b>	8.00	14.00	11.00
<b>korrekt</b>	46	43	89
<b>in %</b>	92.00	86.00	89.00

Total	50	50	100

Dargestellt sind in der ersten Zahlenspalte der Tafel die Fehler bei unverdrehter (unverdr.) Sprechweise (Anzahl = 4, und das ist ein Anteil von 8% an allen Aufgaben = 50) sowie der korrekten Ergebnisse bei unverdrehter Sprechweise (Anzahl = 46, Anteil = 92% an allen Aufgaben = 50), in der zweiten Zahlenspalte die Fehler bei traditionell-verdrehter (verdr.) Sprechweise (Anzahl = 7, Anteil = 14% an allen Aufgaben = 50) sowie der korrekten Ergebnisse bei unverdrehter Sprechweise (Anzahl = 43, Anteil = 86% an allen Aufgaben = 50). Die Zeile bzw. Spalte „Total“ enthält die jeweiligen Summen: z.B.  $4 + 46 = 50$  als eine Spaltensumme und  $4 + 7 = 11$  als eine Zeilensumme. Insgesamt gab es 100 Aufgaben zur Aufgabenstellung „Addition, leicht“.

Zufallskritische Auswertungen der 2x2-Tafeln wurden mit dem csi-Befehl des Statistikpakets Stata14 durchgeführt (Statacorp 2015). Ich ermittelte den Quotienten der Fehleranteile (unverdreht/verdreht), also das Verhältnis der relativen Fehlerhäufigkeiten bei unverdrehter zu verdrehter Sprechweise und zugehörige 95%-Wald-Konfidenzintervalle (Rothman et al. 2008, S. 245-250). Zudem berichte ich zweiseitige p-Werte aus exakten Fisher-Tests (Rothman et al. 2008, S. 255-257; Hartung et al. 2009, S. 416-419). Wenn p-Werte kleiner als 5% sind, nenne ich sie (statistisch) signifikant. Die Ergebnisse dieser Auswertungen werden unter der 2x2-Tafel berichtet:

**Quotient der Fehleranteile (unverdreht/verdreht) =**  
**0.571, 95%-CI: 0.178 - 1.83, 2seitiger p-Wert = 0.525 = 52.5%**

Im Beispiel errechnet sich das Verhältnis der relativen Fehlerhäufigkeiten zu  $8\% / 14\% = 0.571$ . Dieser Quotient von  $0.571 = 57.1\%$  ist die zentrale Effektstatistik. Sie besagt, dass die Fehlerhäufigkeit beim Wechsel auf eine stellenwertgerechte Sprechweise im Mittel um  $100\% - 57.1\% = 42.9\%$  sinkt, d.h. auf 57.1% der Fehlerhäufigkeit bei traditionell-verdrehter Sprechweise. Die Zahl der Fehler ist aber in diesem Beispiel zu gering, um die beobachtete Absenkung bei einem Umfang von je 50 Aufgaben als statistisch signifikant zu sichern: der p-Wert ist mit 52.5% deutlich größer als 5% und das 95%-Konfidenzintervall schließt den Neutralwert des Quotienten ein, der bei 1 liegt, also im Innern des Intervalls von 0.178 bis 1.83.

Gemeinsame Analysen der Tafeln zu unterschiedlichen Aufgabentypen (Variation in Schweregrad und/oder Rechenart) einer Studie erfolgten als kumulierte 2x2-Tafeln mit der oben beschriebenen Methodik sowie mittels logistischer Regression (Rothman et al. 2008, S. 394f; Cameron und Trivedi 2005, S. 466 - 470) unter Berücksichtigung entsprechender Kovariablen wie z.B. dem Schweregrad und/oder der Rechenart. Ich ermittelte zu den ausgewerteten Kovariablen als Effektmaße Chancenverhältnisse („odds ratio“ OR) mit 95%-Wald-Konfidenzintervallen sowie zugehörige z-Tests und zweiseitige p-Werte (Cameron und Trivedi 2010, S. 460-467). Zur Berechnung wurde der `logit`-Befehl aus Stata (Statacorp 2015) verwendet.

Da die Versuche weitgehend balanciert durchgeführt wurden, waren keine relevanten Korrelationen zwischen der Exposition (stellenwertgerechte vs. traditionell-verdrehte Sprechweise) und den Kovariablen (z.B. dem Aufgabentyp) zu erwarten. Deshalb kann ein Confounding weitgehend ausgeschlossen werden. Allerdings ist es von Interesse, statistische Interaktionen zwischen der Exposition und den Kovariablen zu bestimmen, um zu messen, ob die Exposition-Respons-Beziehung durch die Kovariablen modifiziert wird (Rothman et al. 2008, S. 72-75, S. 401-407), also ob z.B. der Einfluss der Sprechweise auf die Fehlerhäufigkeit mit dem Aufgabentyp variiert.

Derartige Auswertungen zu kumulierten Tafeln und die gemeinsamen Regressions-Analysen zu allen Aufgabentypen bleiben lediglich orientierend, da sie die Unabhängigkeit der Beobachtungseinheiten voraussetzen, was aber nicht gegeben ist, denn die Aufgaben aus unterschiedlichen Aufgabentypen wurden von identischen Probandengruppen gelöst. Diese Clusterung konnte ich nicht berücksichtigen, da keine Individualdaten verfügbar waren. Bei gemeinsamer Auswertung von Aufgaben verschiedener Aufgabentypen habe ich deshalb zumindest eine ad-hoc Korrektur zu Konfidenzintervallen und p-Werten angeboten, indem die Tafeleinträge vor der Analyse durch die Zahl der Aufgabentypen geteilt wurden, um so in gewissem Umfang der Verzerrung durch die Clusterung entgegenzuwirken. Die 4-Felder-Einträge wurden nach der Division ganzzahlig gerundet. Dies ad-hoc-Verfahren mag aber in Abhängigkeit von den Umständen der Clusterung sowohl zu einer Unterkorrektur als auch zu einer Überkorrektur führen.

### **3. Schellenberger 1953**

#### **3.1 Versuchsbeschreibung**

Schellenberger war „ein Pionier der Zahlensprechweise“ (Schellenberger 2008). Er führte in den 1950er Jahren Untersuchungen in der DDR durch, um zu ermitteln, ob eine stellenwertgerechte Zahlensprechweise zu einem niedrigeren Fehleranteil führt. Die Untersuchungen wurden vorwiegend mit Absolventen der damals noch existierenden Arbeiter- und Bauernfakultät durchgeführt. Es handelte sich um begabte Schüler, denen durch nachträgliches Ablegen des Abiturs die Möglichkeit des Hochschulstudiums eröffnet werden sollte (Schellenberger 2008).

In einem Vorversuch begann Schellenberger (1953, S. 50f.) mit einem Zahlendiktat in traditionell-verdrehter Zahlensprechweise, bei dem den Probanden zehn neunstellige Zahlen diktiert wurden, wie zum Beispiel 354672785. Danach folgten im Vorversuch zehn Kopfrechenaufgaben zu den vier Grundrechenarten Addition, Subtraktion, Multiplikation und Division, die ebenfalls traditionell-verdreht diktiert wurden. Hier waren jeweils die ersten fünf Aufgaben leichter und die zweiten zehn Aufgaben schwerer gewählt. „Im Anschluß an eine kurze Einführung in die logische Sprechweise der Zahlen und einige Vorübungen wurde dann eine genauso aufgebaute Zahlen- und Aufgabengruppe (natürlich mit anderen Aufgaben) nach der neuen, dem Schriftbild angepaßten Sprechweise angesagt und gerechnet“ (Schellenberger 1953, S. 50).

Schellenberger verwandte eine stellenwertgerechte Sprechweise für Kardinalzahlen, die heute von Zwanzigeins e.V. empfohlen und „zehneins“ (Ze) genannt wird (siehe zur Erläuterung die Ausführungen in Abschnitt 1 „Einleitung“). Die erprobte Zahlensprechweise beschränkte sich nicht auf den Bereich oberhalb von 20. Aber alle zweistelligen Zahlen waren größer als 12, d.h. in stellenwertgerechter Sprechweise wurde z.B. „zehndrei“ bzw. „zehnvier“ diktiert, die Zahlwörter „elf“ und „zwölf“ bzw. „zehneins“ und „zehnzwei“ traten in den Diktaten der Aufgabenstellungen jedoch nicht auf.

„Eine wesentliche Aufgabe der Vorversuche war die Ermittlung der Stoppzeiten“ (Schellenberger 1953, S. 50), die später bei den Hauptversuchen mit Hilfe einer Stoppuhr eingehalten wurden. Eine solche Maximalzeit wurde für jeden Aufgabenblock festgesetzt, aber einheitlich für beide Sprechweisen. Diese in den

Vorversuchen gefundenen Stoppzeiten wurden von Schellenberger aber nicht publiziert.

Der eigentliche Versuch (Schellenberger 1953, S. 51f.), den Schellenberger mit zehn Absolventen der Arbeiter- und Bauernfakultät durchführte, bestand aus vier Aufgabenblöcken zu den vier Grundrechenarten (auf ein Zahlendiktat wurde verzichtet). Auch im Hauptversuch wurden zunächst einfache Kopfrechenaufgaben, dann schwierige gestellt, bei Addition und Multiplikation brachte Schellenberger sogar dreistellige Summanden und Faktoren mit ein (Schwierigkeitsgrad: sehr schwer). Addition und Multiplikation betreffend wurden 15 Aufgaben in der traditionell-verdrehten und 15 Aufgaben in der unverdrehten Zahlensprechweise gestellt, je 5 pro Schwierigkeitsgrad. Bei Subtraktion und Division waren es 10 Aufgaben in jeder Zahlensprechweise, auch jeweils fünf Aufgaben pro Schwierigkeitsgrad. Die Aufgaben wurden diktiert, die Probanden wurden aufgefordert, nur die Ergebnisse zu verschriftlichen. Die Zeit, die die Studenten zum Bearbeiten der einzelnen Aufgaben benötigten, wurde festgehalten, um für jeden Aufgabenblock die Durchschnittszeit zu ermitteln. Nach der im Vorversuch ermittelten Stoppzeit wurde "abgeklopft": wer bis zu diesem Zeitpunkt das Ergebnis nicht ermitteln konnte, hatte einen Strich zu machen.

Alle Zahlen lagen im Bereich von 11 bis 999, wobei glatte Zehner wie 20, 30, 120, 200 usw. ausgeschlossen wurden. Ausnahme: bei den Multiplikationsaufgaben war der Multiplikator einstellig und ebenfalls der Divisor bei Divisionsaufgaben.

Einige Aufgabenbeispiele:  $43+54$  (Addition, leicht),  $74+18$  (Addition, schwer),  $264+532$  (Addition, sehr schwer),  $96-32$  (Subtraktion, leicht),  $85-48$  (Subtraktion, schwer),  $4*17$  (Multiplikation, leicht),  $6*47$  (Multiplikation, schwer),  $7*136$  (Multiplikation, sehr schwer),  $96:3$  (Division, leicht),  $256:4$  (Division, schwer).

Alle im Hauptversuch gestellten Aufgaben sind auf den Seiten 51 und 52 in Schellenberger 1953 gelistet.

Als Zielgröße verfolgte Schellenberger das Verhältnis der Anzahlen korrekter Aufgaben in jedem Aufgabenblock, wobei sich im Zähler die Anzahl bei stellenwertgerechter Sprechweise und im Nenner die Anzahl bei traditionell-verdrehter Sprechweise befand. Im Folgenden wird stattdessen – wie heute üblich – die Zahl der Fehler als Leitgröße verwendet.

### 3.2 Ergebnisse

Die wesentlichen Resultate des Versuchs sind in Tabelle c "Ergebnis" auf S. 52 in Schellenberger 1953 dargestellt, aus denen die folgenden 15 2x2-Tafeln gebildet wurden, die die Grundlage der Sekundäranalyse bilden.

#### 3.2.1 Additionsaufgaben

##### 3.2.1.1 Addition, leicht

	unverdr.	verdr.	Total
-----+-----+-----			
Fehler	4	7	11
in %	8.00	14.00	11.00
-----+-----+-----			
korrekt	46	43	89
in %	92.00	86.00	89.00
-----+-----+-----			
Total	50	50	100

Quotient der Fehleranteile (unverdreh/verdreh) =  
0.571, 95%-CI: 0.178 - 1.83, 2seitiger p-Wert = 0.525 = 52.5%

##### 3.2.1.2 Addition, schwer

	unverdr.	verdr.	Total
-----+-----+-----			
Fehler	9	20	29
in %	18.00	40.00	29.00
-----+-----+-----			
korrekt	41	30	71
in %	82.00	60.00	71.00
-----+-----+-----			
Total	50	50	100

Quotient der Fehleranteile (unverdreht/verdreht) = 0.45, 95%-CI: 0.227 - 0.890, 2seitiger p-Wert = 0.0266 = 2.66%

### 3.2.1.3 Addition, sehr schwer

	unverdr.	verdr.	Total
Fehler	10	18	28
in %	20.00	36.00	28.00
korrekt	40	32	72
in %	80.00	64.00	72.00
Total	50	50	100

Quotient der Fehleranteile (unverdreht/verdreht) = 0.556, 95%-CI: 0.285 - 1.08, 2seitiger p-Wert = 0.118 = 11.8%

### 3.2.1.4 Addition, gepoolt

	unverdr.	verdr.	Total
Fehler	23	45	68
in %	15.33	30.00	22.67
korrekt	127	105	232
in %	84.67	70.00	77.33
Total	150	150	300

Quotient der Fehleranteile (unverdreht/verdreht) = 0.511, 95%-CI: 0.326 - 0.800, 2seitiger p-Wert = 0.0036 = 0.36%

Ad-hoc Korrektur (Division durch 3, gerundet):

Quotient der Fehleranteile (unverdreht/verdreht) =  
 0.533, 95%-CI: 0.249 - 1.14, 2seitiger p-Wert = 0.153 = 15.3%

### 3.2.2 Subtraktionsaufgaben

#### 3.2.2.1 Subtraktion, leicht

	unverdr.	verdr.	Total
<b>Fehler</b>	7	11	18
<b>in %</b>	14.00	22.00	18.00
<b>korrekt</b>	43	39	82
<b>in %</b>	86.00	78.00	82.00
<b>Total</b>	50	50	100

Quotient der Fehleranteile (unverdreht/verdreht) =  
 0.636, 95%-CI: 0.269 - 1.51, 2seitiger p-Wert = 0.436 = 43.6%

#### 3.2.2.2 Subtraktion, schwer

	unverdr.	verdr.	Total
<b>Fehler</b>	6	21	27
<b>in %</b>	12.00	42.00	27.00
<b>korrekt</b>	44	29	73
<b>in %</b>	88.00	58.00	73.00
<b>Total</b>	50	50	100

Quotient der Fehleranteile (unverdreht/verdreht) =  
 0.286, 95%-CI: 0.126 - 0.648, 2seitiger p-Wert = 0.0013 = 0.13%

### 3.2.2.3 Subtraktion, gepoolt

	unverdr.	verdr.	Total
Fehler	13	32	45
in %	13.00	32.00	22.50
korrekt	87	68	155
in %	87.00	68.00	77.50
Total	100	100	200

Quotient der Fehleranteile (unverdreht/verdreht) =  
 0.406, 95%-CI: 0.227 - 0.727, 2seitiger p-Wert = 0.0021 = 0.21%

Ad-hoc Korrektur (Division durch 2, gerundet):

Quotient der Fehleranteile (unverdreht/verdreht) =  
 0.429, 95%-CI: 0.193 - 0.953, 2seitiger p-Wert = 0.525 = 52.5%

### 3.2.3 Multiplikationsaufgaben

#### 3.2.3.1 Multiplikation, leicht

	unverdr.	verdr.	Total
1	0	3	3
	0.00	6.00	3.00
2	50	47	97
	100.00	94.00	97.00

Total		50	50		100
-------	--	----	----	--	-----

Quotient der Fehleranteile (unverdreht/verdreht) = 0.0, 95%-CI: als Wald-Intervall nicht ermittelbar, 2seitiger p-Wert = 0.242 = 24.2%

### 3.2.3.2 Multiplikation, schwer

	unverdr.	verdr.		Total	
-----+-----+-----					
Fehler		25	37		62
in %		50.00	74.00		62.00
-----+-----+-----					
korrekt		25	13		38
in %		50.00	26.00		38.00
-----+-----+-----					
Total		50	50		100

Quotient der Fehleranteile (unverdreht/verdreht) = 0.676, 95%-CI: 0.490 - 0.933, 2seitiger p-Wert = 0.0228 = 2.28%

### 3.2.3.3 Multiplikation, sehr schwer

	unverdr.	verdr.		Total	
-----+-----+-----					
Fehler		16	21		37
in %		32.00	42.00		37.00
-----+-----+-----					
korrekt		34	29		63
in %		68.00	58.00		63.00
-----+-----+-----					
Total		50	50		100

Quotient der Fehleranteile (unverdreht/verdreht) =  
 0.762, 95%-CI: 0.453 - 1.28, 2seitiger p-Wert = 0.408 = 40.8%

### 3.2.3.4 Multiplikation, gepoolt

	unverdr.	verdr.	Total
<b>Fehler</b>	41	61	102
in %	27.33	40.67	34.00
<b>korrekt</b>	109	89	198
in %	72.67	59.33	66.00
<b>Total</b>	150	150	300

Quotient der Fehleranteile (unverdreht/verdreht) =  
 0.672, 95%-CI: 0.486 - 0.930, 2seitiger p-Wert = 0.0204 = 2.04%

Ad-hoc-Korrektur (Division durch 3, gerundet)

Quotient der Fehleranteile (unverdreht/verdreht) =  
 0.70, 95%-CI: 0.400 - 1.22, 2seitiger p-Wert = 0.291 = 29.1%

### 3.2.4 Divisionsaufgaben

#### 3.2.4.1 Division, leicht

	unverdr.	verdr.	Total
<b>Fehler</b>	5	5	10
in %	10.00	10.00	10.00
<b>korrekt</b>	45	45	90
in %	90.00	90.00	90.00

Total		50	50		100
-------	--	----	----	--	-----

Quotient der Fehleranteile (unverdreh/verdreh) =  
 1.00, 95%-CI: 0.309 - 3.24, 2seitiger p-Wert = 1 = 100%

### 3.2.4.2 Division, schwer

	unverdr.	verdr.		Total	
-----+-----+-----					
Fehler		9	16		25
in %		18.00	32.00		25.00
-----+-----+-----					
korrekt		41	34		75
in %		82.00	68.00		75.00
-----+-----+-----					
Total		50	50		100

Quotient der Fehleranteile (unverdreh/verdreh) =  
 0.563, 95%-CI: 0.275 - 1.15, 2seitiger p-Wert = 0.165 = 16.5%

### 3.2.4.3 Division, gepoolt

	unverdr.	verdr.		Total	
-----+-----+-----					
Fehler		14	21		35
in %		14.00	21.00		17.50
-----+-----+-----					
korrekt		86	79		165
in %		86.00	79.00		82.50
-----+-----+-----					
Total		100	100		200

Quotient der Fehleranteile (unverdreh/verdreh) =  
 0.667, 95%-CI: 0.360 - 1.23, 2seitiger p-Wert = 0.264 = 26.4%

Ad-hoc Korrektur (Division durch 2, gerundet)

Quotient der Fehleranteile (unverdreht/verdreht) =  
0.650, 95%-CI: 0.274 - 1.54, 2seitiger p-Wert = 0.437 = 43.7%

### 3.2.5 alle Aufgabentypen, gepoolt

	unverdr.	verdr.	Total
<b>Fehler</b>	91	159	250
in %	18.20	31.80	25.00
<b>korrekt</b>	409	341	750
in %	81.80	68.20	75.00
<b>Total</b>	500	500	1,000

Quotient der Fehleranteile (unverdreht/verdreht) =  
0.572, 95%-CI: 0.457 - 0.717, 2seitiger p-Wert < 0.00005 = 0.005

Ad-hoc Korrektur (Division durch 10, gerundet)

Quotient der Fehleranteile (unverdreht/verdreht) =  
0.563, 95%-CI: 0.275 - 1.15, 2seitiger p-Wert = 0.165 = 16.5%

### 3.2.6 Regressionsanalysen

Regressionsauswertungen belegten die erwartete Abhängigkeit der Fehlerhäufigkeit von der Rechenart (z.B. treten bei Multiplikationsaufgaben häufiger Fehler auf als bei Additionsaufgaben) und dem Schweregrad (so treten häufiger Fehler bei schweren als bei leichten Aufgaben auf), und dies unabhängig von der eingesetzten Zahlensprechweise. Die Interaktionsmodelle ergaben aber keinen Hinweis auf eine modifizierende Wirkung der Rechenart und/oder des Schweregrades der Aufgaben auf den Effekt der stellenwertgerechten Sprechweise, die Fehlerhäufigkeit zu verringern. Die durchschnittliche Schätzung einer Absenkung der Fehlerhäufigkeit um ca. 40% gilt somit bis auf Zufallsfehler für alle Versuchsbedingungen.

### 3.3 Kommentar

Die Ergebnisse zeigten Vorteile einer stellenwertgerechten Sprechweise, denn die Fehlerrate sank im Durchschnitt über alle gestellten Aufgaben um 43 %. Zudem ergab sich in allen Versuchen eine geringere Fehlerrate bei Verwendung einer stellenwertgerechten Sprechweise (einzige Ausnahme: leichte Divisionsaufgaben, wo die Fehlerrate konstant blieb). Die Variation der Ergebnisse zwischen den unterschiedlichen Aufgabentypen war aber statistisch nicht zu sichern, d.h. die Gewinne durch den Einsatz der unverdrehten Sprechweise ergaben sich bei Berücksichtigung der in den Einzeluntersuchungen z.T. ungenügenden Stichprobenumfänge weitgehend einheitlich. Auf dem 5%-Niveau signifikante Absenkungen der Fehlerraten bei Verwendung einer stellenwertgerechten Zahlensprechweise fanden sich allein bei den schweren Aufgabentypen. Bei „Addition, schwer“ sank der Fehleranteil auf 45% (95%-CI: 23% bis 89%,  $p = 2.7\%$ ), bei „Subtraktion, schwer“ auf 29% (95%-CI: 13% bis 65%,  $p = 0.13\%$ ) sowie bei „Multiplikation, schwer“ auf 68% (95%-CI: 49% bis 93%,  $p = 2.3\%$ ).

Schellenberger sah in den Befunden „eine geradezu drastische Illustration für die am Eingang der Arbeit [von ihm] aufgestellte Behauptung ... :

- a) *Unsere Rechenfähigkeit ist viel größer, als es nach unserer Rechenfertigkeit scheinen will,*
- b) *die mangelnde Rechenfertigkeit ist eine Folge der unlogischen Zahlensprechweise in der deutschen Sprache“* (Schellenberger 1953, S. 53f).

Die untersuchten Jugendlichen standen am Ende ihrer Schullaufbahn und die einfachen Aufgabenstellungen konnten keine signifikanten Ergebnisse erbringen, da die Fehlerzahl bereits bei traditionell-verdrehter Sprechweise niedrig war. Um die beobachtete Absenkung der Fehlerquote durch Einsatz einer stellenwertgerechten Sprechweise auch für die einfachen Aufgaben statistisch sichern zu können, hätte der Umfang der Untersuchungen gesteigert werden müssen.

Wie in der Einleitung dargestellt, konnte die Clusterung der Aufgaben pro Schüler in der Auswertung nicht berücksichtigt werden. Insofern mögen die ermittelten Konfidenzintervalle zu eng und die angegebenen p-Werte zu niedrig liegen.

Schellenberger begann die Aufgabenstellungen stets in traditionell-verdrehter Sprechweise und wechselte danach zur unverdrehten Zahlensprechweise. Da also

die Reihenfolge der Sprechweisen nicht variiert wurde, ist eine Überlagerung der Ergebnisse mit Lern- und Carry-over-Effekten (Armitage und Berry 1988, S. 225; Neter et al. 1985, S. 1062) nicht auszuschließen. Diese Effekte wären Verzerrungen zu Gunsten der stellenwertgerechten Sprechweise.

Zehner 1955 berichtete auf S. 14 seiner Habilitationsschrift, dass nach Abschluss seiner Arbeit Schellenberger weitere Ergebnisse publizierte, die – nach Zehner – in die gleiche Richtung weisen (Schellenberger M: Die Angleichung des Zahlworts an das Schriftbild der Zahl. Pädagogik 1955/8, Berlin S. 598). Diese Arbeit war allerdings nicht erreichbar und konnte deshalb in dieser Sekundäranalyse nicht berücksichtigt werden.

Schellenberger beschrieb als eine wesentliche Limitation seiner Befunde: „die Versuchspersonen standen selbstverständlich noch sehr stark unter der Wirkung der von Kindheit her gewohnten Sprechweise.... Ihnen fehlte jede Übung im Rechnen unter Verwendung der logischen Sprechweise“ (Schellenberger 1953, S. 54). Schellenberger folgerte, dass Untersuchungen indiziert seien, die zunächst über längere Zeit hinweg und mit vielen Übungen Schulkinder in die stellenwertgerechte Zahlensprechweise einführen und dann diese so unterrichteten Kinder auf ihre Leistungsfähigkeit testen.

#### **4. Zehner 1955**

Im Nachgang zu den Arbeiten von Schellenberger 1953 beschäftigte sich Kurt Zehner im Rahmen seiner Habilitationsschrift „Das invertierte Zahlensprechen als pädagogisch-psychologisches Problem“ (1955) weitergehend mit der verdrehten Zahlensprechweise.

Auf S. 12 beschreibt der Autor vier prinzipielle Möglichkeiten auf den Widerspruch zwischen Schriftbild und Sprechweise zu reagieren:

- 1) Es bleibt wie es ist, also widersprüchlich.
- 2) Änderung der Ziffernotation: Schreiben von links nach rechts, aber erst den Einer, dann den Zehner schreiben, um so eine Übereinstimmung mit dem Sprechen zu erreichen. D.h. 98 bedeutet neunundachtzig, also nicht 100-2, sondern 100-11.
- 3) Inverses Schreiben der Ziffern: Schreiben der Ziffern von rechts nach links (mit dem vorherigen Setzen von Lücken an den richtigen Stellen), um so eine

Übereinstimmung mit dem Sprechen zu erreichen.

4) Schellenbergers Vorschlag: reformiertes Sprechen, d.h. erst Zehner, dann Einer sprechen, um dadurch eine Übereinstimmung mit dem Schreiben zu erreichen.

Zehner verfolgte Ansatz 4) nach Schellenberger 1953. „Zunächst erscheint es uns erforderlich, diese Versuche SCHELLENBERGERS auszuweiten und mit Schülern in einer natürlichen Schulsituation über einen längeren Zeitraum hinweg systematisch durchzuführen.“ Tatsächlich knüpfte Zehner hier unmittelbar an Schellenbergers Diskussion seiner Ergebnisse an: „Ein echtes Ergebnis, das nach meiner Auffassung zu einem durchschlagenden Erfolg für die logisch richtige Sprechweise führen muß, läßt sich nur erreichen, wenn über eine längere Zeit hinweg Versuche und Übungen in Schulklassen angestellt werden“ (Schellenberger 1953, S. 54).

#### **4.1 Versuchsbeschreibung**

Die Untersuchung bezieht sich auf zwei Klassen aus zwei verschiedenen Potsdamer Grundschulen (Mädchenschulen aus verschiedenen Stadtteilen). Beide Klassen bestanden aus je 27 Mädchen im Alter zwischen 7 und 8 Jahren und „waren zu Beginn der Untersuchung in der Lehrplannerfüllung gleich weit fortgeschritten und wurden auch weiterhin nach dem amtlichen Lehrplan unterrichtet... ein Vergleich des Leistungsstandes erlaubte, das Niveau beider Klassen als ungefähr gleichwertig anzusehen“ (Zehner 1955, S. 17).

Variable „Klasse“:

K=1: Versuchsklasse mit (zunächst) 5 Monaten stellenwertgerechter Sprechweise,

K=0: Vergleichsklasse mit (zunächst) 5 Monaten verdrehter Sprechweise.

Zehner verwendete als Aufgabe ein Zahlendiktat. In der Versuchsklasse (K=1) wurden die Zahlen nicht-invertiert, in der Vergleichsklasse (K=0) invertiert diktiert. Zunächst wurde in einem Vorversuch erhoben, welche durchschnittliche Zeit die Schulkinder benötigten, um eine zweistellige Zahl im Zahlendiktat niederzuschreiben. Es wurde ermittelt, dass 4 Sekunden ausreichten. Das Aufschreiben der Zahlen im späteren Hauptversuch sollte unter einem gewissen Zeitdruck stattfinden. Der zeitliche Abstand zwischen den einzeln diktierten Zahlen sollte ausreichen, um die diktierte Zahl niederzuschreiben, aber nicht, um lange zu überlegen, wie genau das betreffende Zahlbild, d.h. die Ziffernzahl auszusehen hat.

Zehner (1955, S. 16) verwendete eine stellenwertgerechte Sprechweise der Kardinalzahlen, die der von Zwanzigeins e.V. präferierten Sprechweise entspricht („zehneins“, siehe Abschnitt 1 „Einleitung“ zur Erläuterung). Zehner folgte damit dem Ansatz von Schellenberger 1953, aber ging insofern konsequent darüber hinaus, da er auch „zehneins“ und „zehnzwei“ als Zahlwörter für 11 und 12 in den Aufgabenstellungen verwendete.

Die Eltern wurden einbezogen und im Fall der Versuchsklasse gebeten, sich unterstützend einzubringen, aber sich auf keinen Fall während der Versuchsphase gegenüber den Kindern negativ zu der stellenwertgerechten Zahlensprechweise zu äußern.

Der Hauptversuch in Zehner 1955 bestand also aus Zahlendiktaten. Die diktierten Zahlen mussten von den Kindern unmittelbar in indo-arabischer Ziffernform niedergeschrieben werden. In einem Zeitraum von 1 Minute wurden 15 zweistellige Zahlen (ohne volle Zehner) diktiert. Zusätzlich fanden Versuche mit erhöhtem Tempo statt, wo 20 Zahlen pro Minute diktiert wurden. Jede zum Diktat der 15 Zahlen bzw. 20 Zahlen pro Kind entstandene Niederschrift wird im Folgenden als „Arbeit“ bezeichnet. Das gewählte Tempo wird im Folgenden durch eine Geschwindigkeitsvariable  $G$  dargestellt:  $G = 15$  oder  $G = 20$  (Diktat von 15 zweiziffrigen Zahlen pro Minute oder 20 zweiziffrigen Zahlen pro Minute).

Der Versuch wurde in jeder Klasse viermal durchgeführt, so dass  $27 \cdot 4 = 108$  Arbeiten pro Klasse und Versuchsbedingung angestrebt wurden. Diese Wiederholungen sind in den Daten nicht hinreichend dokumentiert, nur die Gesamtergebnisse aus allen 4 Durchgängen je Versuchsbedingung und Klasse sind verfügbar. Zwischen den einzelnen Versuchen lag der zeitliche Abstand von einer Woche. In beiden Klassen wurden dieselben Zahlen benutzt, nur die Zahlensprechweise unterschied sich.

Zusätzlich versuchte Zehner nach 5 Monaten in den beiden Klassen ein Cross-Over durchzuführen, d.h. eine Umstellung in der Sprechweise, wodurch es in jeder Klasse zwei Unterrichtsphasen (Perioden) gab: zunächst 5 Monate stellenwertgerechte Sprechweise in der Versuchsklasse und 5 Monate invertierte Sprechweise in der Vergleichsklasse, dann Umstellung auf verdrehte Sprechweise in der Versuchsklasse und auf die stellenwertgerechte in der Vergleichsklasse.

Darstellung durch eine Umstellungsvariable:  $U = 0$  oder  $1$  beschreibt das "Cross-Over".  $U=0$ : vor Umstellung der Klassen,  $U=1$ : nach Umstellung der Klassen, d.h.  $U=1$  und  $K=1$  ist z.B. die Versuchsklasse nach Umstellung. Diese Cross-Over-Untersuchungen wurden nur mit  $G=20$  durchgeführt.

Daraus ergibt sich als Beschreibung der Expositionsbedingungen (= Sprechweise, "Treatment"):

$E=0$  (traditionell-verdreht: "einundzwanzig"), wenn ( $K=0$  und  $U=0$ ) oder ( $K=1$  und  $U=1$ ), d.h. wenn  $K=U$ ;

$E=1$  (stellenwertgerecht: "zehneins"), wenn ( $K=1$  und  $U=0$ ) oder ( $K=0$  und  $U=1$ ), d.h. wenn  $K \neq U$ .

Zu den Kovariablen:  $K$  und  $U$  legen also  $E$  exakt fest, weshalb  $K$  und  $U$  nicht als Kovariable gleichzeitig in der Auswertung auftreten können. In der Regressionsanalyse wird  $K$  als Kovariable gewählt. Dann gilt:  $E=0, K=0$  so  $U=0$ ;  $E=0, K=1$  so  $U=1$ ;  $E=1, K=0$  so  $U=1$ ;  $E=1, K=1$  so  $U=0$ ; d.h.  $E=K$  bedeutet  $U=0$ ,  $E \neq K$  bedeutet  $U=1$ .

Alle gestellten Aufgaben sind ab S. 171 in Zehner 1955 gelistet. 15 bzw. 20 Aufgaben bilden eine Arbeit, je nach vorgegebenem Tempo ( $G=15, G=20$ ).

Die Ergebnisse wurden wie folgt dokumentiert (Fehlerzahlen):

$N_0$ : Zahl der Arbeiten mit 0 Fehlern

$N_1$ : Zahl der Arbeiten mit 1 Fehler

...

$N_{20}$ : Zahl der Arbeiten mit 20 Fehlern.

Für die Analyse wird jede Arbeit als positiv oder negativ bewertet. Positive Arbeiten haben einen höheren Fehleranteil als negative Arbeiten. Diese Bewertung wird über Cutpoints zur Fehlerzahl in einer Arbeit definiert.

Cutpoint 1: 0% Fehler (kein Fehler);

Cutpoint 2: 20% Fehler (d.h. 3 Fehler bei  $G=15$ , 4 Fehler bei  $G=20$ ).

Positive Arbeit = in der Arbeit liegen mehr Fehler vor als der jeweilige Cutpoint angibt;

Negative Arbeit = in der Arbeit liegen weniger Fehler oder gleich viele Fehler vor, verglichen mit diesem Cutpoint.

Zielgröße ist das Verhältnis der Anteile positiver Arbeiten bei jedem Cutpoint, d.h. es wird der Anteil positiver Arbeiten bei stellenwertgerechter Sprechweise (Zähler) geteilt durch den Anteil der positiven Arbeiten bei traditionell-verdrehter Sprechweise (Nenner).

Es werden drei Vergleiche durchgeführt. Zwei Quervergleiche der Klassen  $K=1$  (Versuchsklasse, unverdrehte Sprechweise) vs.  $K=0$  (Vergleichsklasse, traditionell-verdrehte Sprechweise) bei Tempo  $G=15$  (15 Aufgaben in 1 Minute) bzw.  $G=20$  (20 Aufgaben in 1 Minute). Zudem ein Längsvergleich innerhalb der Versuchsklasse  $K=1$  bei Tempo  $G=20$ .

## 4.2 Ergebnisse

Die zentralen Ergebnisse zum Hauptversuch finden sich in Zehner 1955, S. 27, Abb. 4 ( $U=0$ ,  $G=15$  und  $G=20$ :  $Exp=0$  und  $Exp=1$ ) und S. 44, Abb. 11 ( $U=1$ ,  $G=20$ :  $Exp=0$  und  $Exp=1$ ).

In Abb. 11b (Vergleichsklasse,  $K=0$ ) sind die Kurven vertauscht beschriftet (die Daten wurden korrigiert übernommen).

### 4.2.1 Umfangsübersicht

#### 4.2.1.1 realisierte Umfänge, insgesamt pro Klasse

**Beobachtungseinheit: Arbeit**

K	Anzahl	in %
0	308	50.41
1	303	49.59

<b>Total</b>		<b>611</b>		<b>100.00</b>
--------------	--	------------	--	---------------

#### 4.2.1.2 realisierte Umfänge, pro Klasse im Versuch mit G=15 (nur vor der Umstellung)

Beobachtungseinheit: Arbeit

<b>K</b>		<b>Anzahl</b>		<b>in %</b>
-----+-----				
<b>0</b>		<b>105</b>		<b>51.22</b>
<b>1</b>		<b>100</b>		<b>48.78</b>
-----+-----				
<b>Total</b>		<b>205</b>		<b>100.00</b>

#### 4.2.1.3 realisierte Umfänge, pro Klasse im Versuch mit G=20, ohne und mit Umstellung

Beobachtungseinheit: Arbeit

		<b>U=0</b>		<b>U=1</b>		<b>Total</b>
-----+-----+-----						
<b>K=0</b>		<b>103</b>		<b>100</b>		<b>203</b>
<b>K=1</b>		<b>105</b>		<b>98</b>		<b>203</b>
-----+-----+-----						
<b>Total</b>		<b>208</b>		<b>198</b>		<b>406</b>

### 4.2.2 Analyse 2x2 - Tafeln

#### 4.2.2.1 Cutpoint 1: Zahl der Arbeiten mit mehr als 0 Fehlern

#### 4.2.2.1.1 Cutpoint 1 (0 Fehler), Quervergleich K=1 vs. K=0, G=15

Beobachtungseinheit: Arbeit

Arbeit positiv: mit Fehlern, Arbeit negativ: fehlerfrei

	unverdr.	verdr.	Total
positiv	23	44	67
in %	23.00	41.90	32.68
negativ	77	61	138
in %	77.00	58.10	67.32
Total	100	105	205

Quotient der Anteile positiver Arbeiten (unverdrehet/verdrehet) = 0.549, 95%-CI: 0.360 - 0.838, 2seitiger p-Wert = 0.0046 = 0.46%

#### 4.2.2.1.2 Cutpoint 1 (0 Fehler), Quervergleich K=1 vs. K=0, G=20

Beobachtungseinheit: Arbeit

Arbeit positiv: mit Fehlern, Arbeit negativ: fehlerfrei

	unverdr.	verdr.	Total
positiv	39	73	112
in %	37.14	70.87	53.85
negativ	66	30	96

in %	62.86	29.13	46.15
------	-------	-------	-------

-----+-----+-----

Total	105	103	208
-------	-----	-----	-----

Quotient der Anteile positiver Arbeiten (unverdreht/verdreht) =  
0.524, 95%-CI: 0.397 - 0.692, 2seitiger p-Wert < 0.00005 = 0.005%

#### 4.2.2.1.3 Cutpoint 1 (0 Fehler), Längsvergleich Versuchsgruppe K=1, G=20

Beobachtungseinheit: Arbeit

Arbeit positiv: mit Fehlern, Arbeit negativ: fehlerfrei

	unverdr.	verdr.	Total
--	----------	--------	-------

-----+-----+-----

positiv	39	68	107
---------	----	----	-----

in %	37.14	69.39	52.71
------	-------	-------	-------

-----+-----+-----

negativ	66	30	96
---------	----	----	----

in %	62.86	30.61	47.29
------	-------	-------	-------

-----+-----+-----

Total	105	98	203
-------	-----	----	-----

Quotient der Anteile positiver Arbeiten (unverdreht/verdreht) =  
0.535, 95%-CI: 0.404 - 0.709, 2seitiger p-Wert < 0.00005 = 0.005%

#### 4.2.2.2 Cutpoint 2: Zahl der Arbeiten mit mehr als 20% Fehler

##### 4.2.2.2.1 Cutpoint 2 (20% Fehler), Quervergleich K=1 vs. K=0, G=15

Beobachtungseinheit: Arbeit

Arbeit positiv: mehr als 20% der Aufgaben in der Arbeit falsch,

Arbeit negativ: höchstens 20% der Aufgaben in der Arbeit falsch

	unverdr.	verdr.	Total
positiv	6	19	25
in %	6.00	18.10	12.20
negativ	94	86	180
in %	94.00	81.90	87.80
Total	100	105	205

Quotient der Anteile positiver Arbeiten (unverdrehet/verdrehet) = 0.332, 95%-CI: 0.138 - 0.796, 2seitiger p-Wert = 0.0099= 0.99%

#### 4.2.2.2.2 Cutpoint 2 (20% Fehler), Quervergleich K=1 vs. K=0, G=20

Beobachtungseinheit: Arbeit

Arbeit positiv: mehr als 20% der Aufgaben in der Arbeit falsch,

Arbeit negativ: höchstens 20% der Aufgaben in der Arbeit falsch

	unverdr.	verdr.	Total
positiv	13	39	52
in %	12.38	37.86	25.00
negativ	92	64	156

in %	87.62	62.14	75.00
------	-------	-------	-------

-----+-----+-----

Total	105	103	208
-------	-----	-----	-----

Quotient der Anteile positiver Arbeiten (unverdreht/verdreht) =  
0.327, 95%-CI: 0.186 - 0.576, 2seitiger p-Wert < 0.00005 = 0.005%

#### 4.2.2.2.3 Cutpoint 2 (20% Fehler), Längsvergleich Versuchsgruppe K=1, G=20

Beobachtungseinheit: Arbeit

Arbeit positiv: mehr als 20% der Aufgaben in der Arbeit falsch,

Arbeit negativ: höchstens 20% der Aufgaben in der Arbeit falsch

	unverdr.	verdr.	Total
--	----------	--------	-------

-----+-----+-----

positiv	13	27	40
---------	----	----	----

in %	12.38	27.55	19.70
------	-------	-------	-------

-----+-----+-----

negativ	92	71	163
---------	----	----	-----

in %	87.62	72.45	80.30
------	-------	-------	-------

-----+-----+-----

Total	105	98	203
-------	-----	----	-----

Quotient der Anteile positiver Arbeiten (unverdreht/verdreht) =  
0.449, 95%-CI: 0.246 - 0.820, 2seitiger p-Wert = 0.0080 = 0.80%

#### 4.2.3 Regressionen

Die Regressionsanalysen zeigten ein ähnliches Ausgangsniveau in der Fehlerzahl bei Einsatz der traditionell-verdrehten Sprechweise für beide Klassen, aber eine Interaktion der Sprechweise mit der Klasse: die Versuchsgruppe (K=1) zeigte bei

Einsatz der stellenwertgerechten Sprechweise eine bessere Leistung als die Vergleichsklasse ( $K=0$ ). Diese Beobachtung wird im Kommentar 4.3 vertieft, wo die Belastbarkeit der Umstellung in der Vergleichsklasse ( $K=0$ ) und das Ausgangsniveau beider Klassen näher betrachtet werden. Wie erwartet, zeigten die Regressionsanalysen eine Abhängigkeit der Fehlerzahl von der Diktatgeschwindigkeit mit deutlich häufigeren Fehlern bei  $G=20$  als bei  $G=15$ . Allerdings ergab sich kein Hinweis auf eine modifizierende Wirkung von  $G$ , sobald die Klasse als Kovariable berücksichtigt wurde. Diese Ergebnisse fanden sich sowohl bei Verwendung des Cutpoints 1 (= Fehler) als auch bei Verwendung des Cutpoints 2 (20% Fehler).

### 4.3 Kommentar

Die Zahl der Arbeiten mit Fehlern war in der Versuchsklasse niedriger: um 45% bei 15 diktierten Zahlen pro Minute, um 48% bei 20 Zahlen pro Minute; der Unterschied war noch stärker ausgeprägt mit je einer Reduktion der positiven Arbeiten um 67% bei Verwendung des höheren Cutpoints von 20% fehlerhaften Aufgaben in der Arbeit. Der Längsvergleich in der Versuchsklasse 1 zeigte Verringerungen der Zahl an Arbeiten mit Fehlern bei stellenwertgerechter Sprechweise um 47% (15 Zahlen pro Minute) bzw. 55% (20 Zahlen pro Minute). Alle zugehörigen zweiseitigen  $p$ -Werten lagen unter 1%. Die Geschwindigkeit ( $G=15$  vs.  $G=20$ ) modifizierte diese relative Fehlerreduktion durch die stellenwertgerechte Sprechweise kaum: die prozentualen Gewinne waren trotz unterschiedlichem Fehlerniveau bei  $G=15$  und  $G=20$  ähnlich. „In diesem Sinne ergibt sich aus der vorliegenden Arbeit [=Habilitationsschrift] als Folgerung für den Rechenunterricht: Solange im deutschen Sprachraum die Zahlen noch invertiert gesprochen werden, ist die Einführung der nicht-invertierten Zahlensprechweise als methodischer Schritt im zweiten Schuljahr zu empfehlen“ (Zehner 1955, S. 162).

Ein wichtiger Aspekt ist die Querschnittsstruktur beim Vergleich von Versuchs- und Vergleichsklasse. Hier können nicht vermessene Kovariablen, die sich in ihrer Häufigkeitsverteilung zwischen den Klassen unterscheiden, diese Vergleiche stören. Zehner argumentierte zwar, dass nach seiner Einschätzung beide Klassen „zu Beginn der Untersuchung in der Lehrplanerfüllung gleich weit fortgeschritten [waren]

und ... auch weiterhin nach dem amtlichen Lehrplan unterrichtet [wurden]“ und dass „ein Vergleich des Leistungsstandes erlaubte, das Niveau beider Klassen als ungefähr gleichwertig anzusehen“ (Zehner 1955, S. 17). Dies kann aber eine angemessene Prüfung nicht ersetzen. Zumindest erlauben die publizierten Daten eine Gegenüberstellung der Fehlerhäufigkeit in beiden Klassen bei Verwendung der traditionell-verdrehten Sprechweise als Referenzgröße. Für die Versuchsklasse (K=1) fand dies nur bei höherer Diktiergeschwindigkeit (G=20) und nach der Umstellung (U=1) statt, bei der Vergleichsklasse (K=0) mit G=20 vor der Umstellung (U=0). Die nachstehende 2x2-Tafel vergleicht die beiden Klassen K=1 und K=0 unter diesen Bedingungen.

**Beobachtungseinheit: Arbeit**

**Arbeit positiv: mit Fehlern, Arbeit negativ: fehlerfrei**

	K=1	K=0	Total
positiv	68	73	141
in %	69.39	70.87	70.15
negativ	30	30	60
in %	30.61	29.13	29.85
Total	98	103	201

Quotient der Anteile positiver Arbeiten (unverdrehet/verdrehet) = 0.980, 95%-CI: 0.817 - 1.17, 2seitiger p-Wert = 0.878 = 87.8%

Versuchs- (K=1) und Vergleichsklasse (K=0) unterscheiden sich kaum in ihrem Fehlerniveau, wenn die traditionelle Zahlensprechweise im Diktat verwendet wird. Dies spiegelte auch die Regressionsanalyse (siehe 4.2.3). Das Ergebnis stützt die von Zehner berichtete Ähnlichkeit beider Klassen. Zudem ergab sich der Befund des Vorteils einer stellenwertgerechten Sprechweise auch im Längsvergleich innerhalb der Versuchsklasse (K=1), was die Ergebnisse aus dem Quervergleich beider Klassen untermauert.

Von Interesse wäre daher auch eine Längsschnittstudie zur Vergleichsklasse ( $K=0$ ). Die Umstellung ( $U=1$ ) hatte bei der Vergleichsklasse allerdings nicht funktioniert, so dass dieser Teil der Untersuchung nicht belastet werden kann. Siehe hierzu die Erklärungen von Zehner im nachstehenden Textauszug (Zehner 1955, S. 36):

7. Ergebnisse nach Umstellung der Sprechweise  
in beiden Klassen

Nachdem wir die Versuche zunächst abgeschlossen hatten, stellten wir die nunmehr fünf Monate nicht-invertiert sprechende Vukl. auf die übliche Sprechweise um, was ohne Schwierigkeiten vor sich ging und setzten alsdann die Untersuchungen fort.<sup>1)</sup>

Gleichzeitig führten wir die neue - nicht-invertierte - Sprechweise in gewissem Umfang in der Vgkl. ein. Das geschah in der Art, daß wir jede Rechenstunde, in der eine Kontrollarbeit geschrieben wurde, nach entsprechender Vorbereitung in der neuen S. rechweise abhielten. Dieses Vorgehen ist natürlich nicht völlig einwandfrei und den hiermit gewonnenen Ergebnissen wird nur ein bedingter Aussagewert zuzumessen sein. Es war aber leider nicht möglich, die Vgkl. für einen längeren Zeitraum dauernd umzustellen und auch dann hätten wir keine vollkommen einwandfreie Vergleichsmöglichkeit mit der Vukl. gehabt, da die Vgkl. doch schon während der ersten fünf Monate des Schuljahres invertiert gesprochen hatte.

Insofern ist das geplante Cross-Over der Versuchsbedingungen (Arten der Zahlensprechweise) nicht gelungen. Da also die Reihenfolge der Sprechweisen in der verbliebenen Längsschnittuntersuchung nicht variiert wurde, ist eine Überlagerung der Ergebnisse mit Lern- und Carry-over-Effekten (Armitage und Berry 1988, S. 225; Neter et al. 1985, S. 1062) nicht auszuschließen. Abb. 3 auf S. 26 in Zehner 1955 zeigt eine Leistungssteigerung in beiden Klassen während der 4 Wiederholungsmessungen mit Abstand von je 1 Woche, so dass Lern- und Carry-over-Effekte von Bedeutung sein können. Es standen aber nur Prozent- und keine Absolutzahlen zu Abb. 3 zur Verfügung, so dass zu diesen Daten keine Sekundäranalyse möglich war.

Wie in der Einleitung dargestellt, konnte die Clusterung der Aufgaben pro Schüler in der Auswertung nicht berücksichtigt werden (es gab 4 Wiederholungen der Tests je Schüler und Untersuchungsbedingung). Insofern mögen die ermittelten Konfidenzintervalle zu eng und die angegebenen p-Werte zu niedrig liegen.

Zehner 1955 hat weitere Untersuchungen neben dem hier sekundär analysierten Zahlendiktaten im Rahmen seiner Habilitationsarbeit durchgeführt:

- a) In den beiden Mädchenklassen ( $K=1$  und  $K=0$ ) führte Zehner besondere „Tempoversuche“ durch. Zehner begann mit 10 diktierten Zahlen pro Minute und endete mit 37 diktierten Zahlen pro Minute. Auch hier zeigte sich die Überlegenheit der Versuchsklasse, in der mit stellenwertgerechter Sprechweise diktiert wurde.
- b) In den beiden Mädchenklassen ( $K=1$  und  $K=0$ ) erfolgten zudem sog. „Reproduktionsversuche“, ebenfalls bestehend aus Zahlendiktaten. Zweistellige Zahlen wurden in dieser Versuchsreihe unterschiedlich präsentiert (optische, akustische und optisch-akustische Darbietung), und die Fehlerhäufigkeit in den beiden Klassen wurde im Querschnitt verglichen, d.h. Vergleich der Wirkung der beiden Sprechweisen in Abhängigkeit von der Präsentationsform. Dabei wurden die Ergebnisse sowohl beim direkten Niederschreiben der präsentierten Zahlen als auch bei verlangter Reproduktion der Zahlen nach 10 min, nach 30 min und nach 1 Woche ermittelt. Auch hier beschrieb Zehner einen Vorteil durch den Einsatz der stellenwertgerechten Sprechweise und diskutierte die Befunde unter psychologischen Aspekten.
- c) Des Weiteren unternahm Zehner in den beiden Mädchenklassen ( $K=1$  und  $K=0$ ) Versuche zur Fehlerzahl bei Aufgaben mit Rechenoperationen (Addition in zwei Schweregraden, Subtraktion). Wiederum ergab sich in der Versuchsklasse mit stellenwertgerechter Sprechweise ein besserer Befund.
- d) Zehner beschäftigte sich intensiv mit der Häufigkeit des Auftretens von Inversionsfehlern. Es zeigte sich, dass die Leistungsnachteile bei verdrehter Sprechweise in der Vergleichsklasse  $K=0$  mit einer erhöhten Zahl an Inversionsfehlern einherging.
- e) Zehner führte zusätzlich zu den Studien in den beiden Mädchenklassen ( $K=1$  und  $K=0$ ) aus Potsdamer Grundschulen eine Untersuchung mit Umstellung der Zahlensprechweise in einer Sonderschule durch. Die Ergebnisse waren

auch aufgrund des geringeren Umfangs (8 Schulkinder) weniger klar als in den Grundschulen, und Zehners Bericht konzentrierte sich auf kasuistische Darstellungen. Es ergaben sich eindeutige Vorteile der stellenwertgerechten Sprechweise bei vier Kindern, eine uneinheitliche Situation bei 3 Kindern und eine Verschlechterung bei einem der Kinder.

## **5. Ammareller 2006 (Schüler)**

### **5.1 Versuchsbeschreibung**

Die Untersuchungen (Zahlendiktate und Kopfrechenaufgaben) fanden in zwei Klassen (Klassen 1 und 2) der zweiten Stufe einer Grundschule in Bochum statt. Die Flurbezeichnung dieser Grundschule ist „Waldschule“, siehe <https://www.waldschule-bochum.de/> (siehe zur Beschreibung der Schule auch S. 54 – 59 in Ammareller 2006). An dieser Grundschule wurde probeweise die stellenwertgerechte Zahlensprechweise systematisch gelehrt (Ammareller 2008, Eiskirch 2008), wobei auch Zahlen im Zahlbereich unter 20 unverdreht gesprochen wurden (Ammareller 2006, S. 61). „Seit dem Schuljahr 2005/2006 ist die Anwendung der unverdrehten Zahlensprechweise per Schulkonferenz für die Waldschule beschlossen worden, in Absprache mit den Elternvertretern. Das jetzige zweite Schuljahr war die erste Klasse, die von Schulbeginn an mit der `neuen` Zahlensprechweise in Berührung kam“ (Ammareller 2006, S. 60). „Es ist zu betonen, dass das unverdrehte Zahlensprechen an der Waldschule kein Zwang ist. Es bleibt den Schülern überlassen, in welcher Zahlensprechweise sie sich äußern... Die Konsequenz daraus ist, dass die Schüler beide Zahlensprechweisen gut beherrschen, da sie mit beiden Zahlensprechweisen im Unterricht umgehen“ (Ammareller 2006, S. 61). Die Umsetzung der unverdrehten Sprechweise im Mathematikunterricht an der Waldschule wurde am Beispiel einer Unterrichtsstunde detailliert erläutert (Ammareller 2006, S. 62 – 70).

Aus Klasse 1 nahmen 22 Kinder an der Untersuchung teil, darunter 13 Mädchen und 9 Jungen. Zehn der Kinder waren fremdsprachig. Der Mathematikunterricht wurde durch Frau E. erteilt. In Klasse 2 wurde der Versuch mit 18 Kindern durchgeführt, davon 6 Mädchen und 12 Jungen. Nur eine Schülerin war deutschsprachig, alle anderen Kinder fremdsprachig. Den Mathematikunterricht erteilte Frau St. Beide

Klassen waren in der Multiplikation noch nicht geübt. Klasse 2 nahm nicht an der Multiplikationsaufgabe teil. Beiden Klassen wurden ansonsten dieselben Aufgaben gestellt.

Vier Aufgabentypen wurden zunächst in stellenwertgerechter und dann in verdrehter Sprechweise mit anderen, aber ähnlichen Zahlen präsentiert: je 8 diktierte Zahlen im Zahlenbereich bis 100; 8 Additionsaufgaben im Zahlenbereich bis 20; 8 Subtraktionsaufgaben im Zahlenbereich bis 20, jedoch ohne Zehnerunterschreitung; 3 Multiplikationsaufgaben als Verdopplungsaufgaben (nur eine Aufgabe enthielt einen zweistelligen Faktor). Alle Aufgaben sind auf den Seiten 94 und 95 in Ammareller 2006 zusammengestellt.

Die angestrebten Aufgabenumfänge pro Klasse umfassten also für die ersten drei Aufgabentypen in Klasse 1 je Typ  $22 \cdot 8 \cdot 2 = 352$  Aufgaben und in Klasse 2 je Typ  $18 \cdot 8 \cdot 2 = 288$  Aufgaben; für den Typ „Multiplikationsaufgabe“ in Klasse 1 zusätzlich  $22 \cdot 3 \cdot 2 = 132$  Aufgaben. Mit Ausnahme der Subtraktionsaufgaben in Klasse 2 (nur 272 auswertbare Aufgaben) konnten alle angestrebten Umfänge realisiert werden.

„Die Schüler wurden dazu aufgefordert, nur die Ergebnisse der Rechenaufgaben aufzuschreiben und nicht die Rechenaufgaben selbst. Es wurde betont, dass eventuelle Fehler keine Konsequenzen hätten, sondern es nur darum ginge, dass jeder rechne wie er könne. Es wurden vorbereitete Blätter ausgeteilt, auf denen die Schüler ihre Ergebnisse schriftlich festhielten“ (Ammareller 2008, S. 68). Die Schüler der Klasse 1 wurden auf die Möglichkeit hingewiesen, „für eine zu schwierig empfundene Aufgabe einen `Strich` zu machen“ (Ammareller 2008, S. 105).

Im Folgenden werden die Anzahl der aufgetretenen Fehler insgesamt (ohne Betrachtung der Art der Fehler) ausgewertet. Vergleiche finden ausschließlich innerhalb der Klassen statt.

## 5.2 Ergebnisse

### 5.2.1 Zahlendiktat

#### 5.2.1.1 Zahlendiktat Klasse 1

unverdr.	verdr.	Total
-----+	-----+	-----

Fehler		16	38		54
in %		9.09	21.59		15.34
-----+-----+-----					
korrekt		160	138		298
in %		90.91	78.41		84.66
-----+-----+-----					
Total		176	176		352

Quotient der Fehleranteile (unverdreh/verdreh) =  
0.421, 95%-CI: 0.244 - 0.726, 2seitiger p-Wert = 0.0017 = 0.17%

### 5.2.1.2 Zahlendiktat Klasse 2

		unverdr.	verdr.		Total
-----+-----+-----					
Fehler		0	23		23
in %		0.00	15.97		7.99
-----+-----+-----					
korrekt		144	121		265
in %		100.00	84.03		92.01
-----+-----+-----					
Total		144	144		288

Quotient der Fehleranteile (unverdreh/verdreh) =  
0.0, 2seitiger p-Wert < 0.00005 = 0.005%

### 5.2.1.3 Zahlendiktat Klasse 1 und 2

		unverdr.	verdr.		Total
-----+-----+-----					
Fehler		16	61		77
in %		5.00	19.06		12.03
-----+-----+-----					
korrekt		304	259		563
in %		95.00	80.94		87.97

	unverdr.	verdr.	Total
Total	320	320	640

Quotient der Fehleranteile (unverdreh/verdreh) =  
0.262, 95%-CI: 0.155 - 0.445, 2seitiger p-Wert < 0.00005 = 0.005%

## 5.2.2 Addition

### 5.2.2.1 Addition Klasse 1

	unverdr.	verdr.	Total
Fehler	47	59	106
in %	26.70	33.52	30.11
korrekt	129	117	246
in %	73.30	66.48	69.89
Total	176	176	352

Quotient der Fehleranteile (unverdreh/verdreh) =  
0.797, 95%-CI: 0.578 - 1.10, 2seitiger p-Wert = 0.201 = 20.1%

### 5.2.2.2 Addition Klasse 2

	unverdr.	verdr.	Total
Fehler	11	19	30
in %	7.64	13.19	10.42
korrekt	133	125	258
in %	92.36	86.81	89.58
Total	144	144	288

Quotient der Fehleranteile (unverdreht/verdreht) =  
 0.579, 95%-CI: 0.286 - 1.173, 2seitiger p-Wert = 0.176 = 17.6%

### 5.2.2.3 Addition Klasse 1 und 2

	unverdr.	verdr.	Total
<b>Fehler</b>	58	78	136
in %	18.13	24.38	21.25
<b>korrekt</b>	262	242	504
in %	81.88	75.63	78.75
<b>Total</b>	320	320	640

Quotient der Fehleranteile (unverdreht/verdreht) =  
 0.744, 95%-CI: 0.550 - 1.007, 2seitiger p-Wert = 0.0661 = 6.61%

### 5.2.3 Subtraktion

#### 5.2.3.1 Subtraktion Klasse 1

	unverdr.	verdr.	Total
<b>Fehler</b>	26	49	75
in %	14.77	27.84	21.31
<b>korrekt</b>	150	127	277
in %	85.23	72.16	78.69
<b>Total</b>	176	176	352

Quotient der Fehleranteile (unverdreht/verdreht) =  
 0.531, 95%-CI: 0.346 - 0.813, 2seitiger p-Wert = 0.0040 = 0.4%

#### 5.3.2.2 Subtraktion Klasse 2

	unverdr.	verdr.	Total
-----+-----+-----			
Fehler	22	25	47
in %	16.18	18.38	17.28
-----+-----+-----			
korrekt	114	111	225
in %	83.82	81.62	82.72
-----+-----+-----			
Total	136	136	272

Quotient der Fehleranteile (unverdreh/verdreh) =  
0.880, 95%-CI: 0.522 - 1.482, 2seitiger p-Wert = 0.749 = 74.9%

### 5.3.2.3 Subtraktion Klasse 1 und 2

	unverdr.	verdr.	Total
-----+-----+-----			
Fehler	48	74	122
in %	15.38	23.72	19.55
-----+-----+-----			
korrekt	264	238	502
in %	84.62	76.28	80.45
-----+-----+-----			
Total	312	312	624

Quotient der Fehleranteile (unverdreh/verdreh) =  
0.649, 95%-CI: 0.467 - 0.900, 2seitiger p-Wert = 0.0114 = 1.14%

### 5.2.4 Multiplikation (nur Klasse 1)

	unverdr.	verdr.	Total
-----+-----+-----			
Fehler	13	19	32

in %	19.70	28.79	24.24
-----+-----+-----			
korrekt	53	47	100
in %	80.30	71.21	75.76
-----+-----+-----			
Total	66	66	132

Quotient der Fehleranteile (unverdreht/verdreht) =  
0.684, 95%-CI: 0.369 - 1.27, 2seitiger p-Wert = 0.309 = 30.9%

### 5.2.5 alle Rechenaufgabentypen (inkl. Multiplikation): Klasse 1 und 2

	unverdr.	verdr.	Total
-----+-----+-----			
Fehler	119	171	290
in %	17.05	24.50	20.77
-----+-----+-----			
korrekt	579	527	1,106
in %	82.95	75.50	79.23
-----+-----+-----			
Total	698	698	1,396

Quotient der Fehleranteile (unverdreht/verdreht) =  
0.70, 95%-CI: 0.56 - 0.86, 2seitiger p-Wert = 0.0007 = 0.07%

### Ad-hoc-Korrektur (Division durch 3, gerundet)

3 Aufgabentypen; 2.5 Aufgabentypen pro Schüler,  
da Multiplikation nur in Klasse 1:

Quotient der Fehleranteile (unverdreht/verdreht) =  
0.702, 95%-CI: 0.489 - 1.01, 2seitiger p-Wert = 0.0676 = 6.76%

### 5.2.6 alle Aufgabentypen (inkl. Multiplikation): Klasse 1 und 2

	unverdr.	verdr.	Total
--	----------	--------	-------

Fehler	135	232	367
in %	13.26	22.79	18.03
korrekt	883	786	1,669
in %	86.74	77.21	81.97
Total	1,018	1,018	2,036

Quotient der Fehleranteile (unverdreht/verdreht) =  
0.582, 95%-CI:0.479 - 0.706, 2seitiger p-Wert < 0.00005 = 0.005%

Ad-hoc-Korrektur (Division durch 4, gerundet)

4 Aufgabentypen; 3.5 Aufgabentypen pro Schüler,  
da Multiplikation nur in Klasse 1:

Quotient der Fehleranteile (unverdreht/verdreht) =  
0.586, 95%-CI:0.398 - 0.862, 2seitiger p-Wert = 0.0078 = 0.78%

## 5.2.6 alle Aufgabentypen: logistische Regression

### 5.2.6.1 logistische Regression mit Kovariable Klasse

logistische Regression mit Kovariable Klasse mit ad-hoc-Korrektur:  
Zähler und Nenner durch 4 geteilt (gerundet), da 3.5 Aufgabentypen pro Schüler.

E=0: traditionell-verdrehte Zahlensprechweise,  
E=1: stellenwertgerechte Sprechweise.

Klasse\_ind=0: Frau E. (Klasse 1),  
Klasse\_ind=1: Frau St. (Klasse 2).

Logistic regression for grouped data                      Number of obs        =            510

_outcome	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
E	.5223149	.1238253	-2.74	0.006	.3281994    .8312411
Klasse_ind	.4676971	.1181152	-3.01	0.003	.2850997    .7672423
_cons	.3975164	.067605	-5.42	0.000	.2848345    .5547759

## 5.2.6.2 logistische Regression mit Interaktion Klasse und Interaktion Aufgabe

logistische Regression mit Kovariablen und Interaktionen für Klasse und Aufgabentyp mit ad-hoc-Korrektur:

Zähler und Nenner durch 4 geteilt (gerundet), da 3.5 Aufgabentypen pro Schüler.

E=0: traditionell-verdrehte Zahlensprechweise,

E=1: stellenwertgerechte Sprechweise.

Klasse\_ind=0: Frau E. (Klasse 1),

Klasse\_ind=1: Frau St. (Klasse 2)

Aufg\_typ=1\_Diktat,

Aufg\_typ=2\_Addition,

Aufg\_typ=3\_Subtraktion,

Aufg\_typ=4\_Multiplikation.

Int: Interaktionsterme mit Sprechweise E.

Logistic regression for grouped data                      Number of obs        =            510

```
-----
```

_____	_____	_____	_____	_____	_____	_____
_outcome	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
-----+	-----	-----	-----	-----	-----	-----
E	.2317008	.0708424	-4.78	0.000	.1272545	.4218732
Klasse_ind	.490599	.081033	-4.31	0.000	.3549216	.6781423
Aufg_typ_2	1.37712	.2682822	1.64	0.100	.9400394	2.017427
Aufg_typ_3	1.314411	.258606	1.39	0.165	.8938405	1.932868
Aufg_typ_4	1.29487	.4047768	0.83	0.408	.7016823	2.389528
Int_Klasse	.8315468	.2275157	-0.67	0.500	.4863997	1.421609
Int_Aufg_2	3.124406	1.108959	3.21	0.001	1.558284	6.264525
Int_Aufg_3	2.643548	.9539435	2.69	0.007	1.303241	5.362279
Int_Aufg_4	2.618691	1.343421	1.88	0.061	.9580918	7.1575
_cons	.3121975	.0481743	-7.54	0.000	.2307188	.4224506

```
-----
```

### 5.3 Kommentar

Bei gemeinsamer Auswertung aller gestellten Aufgaben (Anzahl = 2036) ergab sich eine Fehlerreduktion bei stellenwertgerechter Sprechweise im Vergleich zur traditionell-verdrehten Form im Durchschnitt von ca. 40% (2seitiger p-Wert < 0.005%; auch nach ad-hoc Korrektur signifikant: p = 3.5%). Die Fehlerverringering war im

Zahlendiktat stärker ausgeprägt (Reduktion um ca. 70%) und statistisch hochsignifikant ( $p < 0.005\%$ ), wobei in Klasse 2 nur bei verdrehter Sprechweise überhaupt Fehler im Diktat auftraten. Bei den Rechenaufgaben wurden die Fehler bei Verwendung einer stellenwertgerechten Form im Mittel um ca. 30% verringert (nach ad-hoc Korrektur grenzwertig signifikant:  $p = 6.7\%$ ), um 25% bei den Additionsaufgaben ( $p = 6.6\%$ ) und um 35% bei den Subtraktionsaufgaben ( $p$  ungefähr 1%). Auch bei den Multiplikationsaufgaben fand sich eine Reduktion der Fehler (um ca. 30%), wenn auch deutlich nicht-signifikant (allein Klasse 1 nahm teil,  $p$  bei ca. 30%). „Beide Klassen erzielten in *allen* Versuchsbereichen – mehr oder weniger deutlich – bessere Ergebnisse unter Anwendung der unverdrehten Zahlensprechweise“ (Ammareller 2006, S. 106). Ammareller (2006, S. 110) sieht deshalb den „Vorteil der nicht-invertierten Zahlensprechweise“ durch ihre Untersuchung bestätigt, trotz methodischer Einschränkungen. So ereigneten sich in Klasse 2 im Mittel ungefähr halb so viele Fehler wie in Klasse 1 (Unterschied im Regressionsmodell mit der Klasse als Kovariable auch nach ad-hoc Korrektur statistisch gesichert mit  $p < 0.3\%$ , siehe 5.2.6.1 und 5.2.6.2).

„Es ist zu bemerken, dass die Versuchsergebnisse der beiden Klassen nur eingeschränkt vergleichbar sind:

- Bei den Aufgabenstellungen in Klasse 1 habe ich ausdrücklich betont, dass die Möglichkeit besteht, statt dem Ergebnis auch einen „Strich“ zu ziehen, falls eine Lösung nicht möglich wäre. In Klasse 2 habe ich diese Möglichkeit nicht eingeräumt.

Dieses spiegelt sich in den Ergebnissen wieder. Der Anteil der nicht bearbeiteten Aufgaben in Klasse 1 ist im Vergleich zu Klasse 2 wesentlich höher. Die Schüler in Klasse 2 zogen erst gar nicht in Erwägung, eine Aufgabe nicht zu bearbeiten.

Es wäre sinnvoller gewesen, hier in beiden Klassen eine einheitliche Vorgabe zu machen.

- Es ist anzumerken, dass der Versuch in Klasse 1 in der vierten Stunde durchgeführt wurde und in Klasse 2 in der ersten Stunde. Es muss berücksichtigt werden, dass die Schüler in der ersten Stunde konzentrierter arbeiten konnten als in der vierten Stunde“ (Ammareller 2006, S. 107).

Als Interaktionseffekt lässt sich im logistischen Regressionsmodell nach Anwendung der ad-hoc Korrektur allein Folgendes sichern: im Zahlendiktat ergab sich eine

größere Fehlerreduktion durch die stellenwertgerechte Sprechweise im Vergleich zu den Rechenaufgaben (die Odds Ratios der Interaktionsterme zu den Rechenaufgaben sind größer als 2.5 und deren p-Werte niedrig, vgl. 5.2.6.2).

Wie in der Einleitung dargestellt, konnte die Clusterung der Aufgaben pro Schüler in der Auswertung nicht berücksichtigt werden. Insofern mögen die ermittelten Konfidenzintervalle zu eng und die angegebenen p-Werte zu niedrig liegen.

Ammareller begann die Aufgabenstellungen stets in stellenwertgerechter Sprechweise und wechselte danach zur traditionell-verdrehten Form. Da also die Reihenfolge der Sprechweisen nicht variiert wurde, ist eine Überlagerung der Ergebnisse mit Lern- und Carry-over-Effekten (Armitage und Berry 1988, S. 225; Neter et al. 1985, S. 1062) nicht auszuschließen. Diese Effekte wären Verzerrungen zu Ungunsten der stellenwertgerechten Sprechweise.

„Ich hoffe, dass ich mit meiner vorliegenden Arbeit dazu beitragen kann, auf die Schwierigkeit unserer invertierten Zahlensprechweise aufmerksam zu machen und Sensibilität für dieses Thema zu wecken, vielleicht sogar die Bereitschaft zu fördern, fundierte Untersuchungen auf breiterer Basis durchzuführen, um neue objektivere, repräsentativere Ergebnisse zu diesem Thema zu erzielen“ (Ammareller 2006, S. 111).

## **6. Überblick und Diskussion**

Zu drei empirischen pädagogischen Studien (Schellenberger 1953, Zehner 1955, Ammareller 2006) wurde eine sekundäre statistische Analyse mit Tabellenauswertung und logistischer Regression durchgeführt, in der die Fehlerhäufigkeiten unter Schülern bei stellenwertgerechter und bei traditionell-verdrehter Sprechweise der Zahlen verglichen wurden. In allen drei Arbeiten wurde eine stellenwertgerechte Sprechweise für Kardinalzahlen verwendet, die heute von Zwanzigeins e.V. empfohlen und „zehneins“ (Ze) genannt wird (siehe zur Erläuterung Abschnitt 1 „Einleitung“ und zur Definition der Sprechweise: Zwanzigeins 2025b). Zur Detaildiskussion der drei Studien und ihrer Limitationen sei auf die entsprechenden Kommentarkapitel in diesem Bericht verweisen (Schellenberger 1953: Kapitel 3.3, Zehner 1955: Kapitel 4.3, Ammareller 2006: Kapitel 5.3).

Schellenberger 1953 (Kopfrechenaufgaben): 10 Absolventen der Arbeiter- und Bauernfakultät wurden vier Aufgabenblöcke zu den vier Grundrechenarten unterschiedlicher Schwierigkeitsgrade gestellt (auf ein Zahlendiktat wurde verzichtet). Die Ergebnisse zeigten Vorteile einer stellenwertgerechten Sprechweise, denn die Fehlerrate sank im Durchschnitt über alle gestellten Aufgaben um 43 %. Auf dem 5%-Niveau signifikante Absenkungen der Fehlerraten fanden sich allerdings allein bei den schweren Aufgabentypen.

Zehner 1955 (Zahlendiktate): Zwei Klassen (2 x 27 Mädchen im Alter zwischen 7 und 8 Jahren aus zwei Potsdamer Grundschulen), eine Klasse wurde über fünf Monate in stellenwertgerechter Sprechweise unterrichtet, die andere – wie üblich – in traditionell-verdrehter Sprechweise. Die Zahl der Arbeiten mit Fehlern war bei stellenwertgerechter Zahlensprechweise um ca. 45% bei 15 diktierten Zahlen pro Minute niedriger, um ca. 50% bei 20 diktierten Zahlen pro Minute. Die zugehörigen zweiseitigen p-Werte lagen unter 1%.

Ammareller 2006 (Zahlendiktate und Kopfrechenaufgaben): Zwei Klassen der 2. Stufe einer Grundschule in Bochum (insgesamt 40 Kinder), die in stellenwertgerechter Sprechweise seit dem ersten Schuljahr unterrichtet worden waren. Bei gemeinsamer Auswertung aller gestellten Aufgaben ergab sich eine Fehlerreduktion bei stellenwertgerechter Sprechweise im Vergleich zur traditionell-verdrehten Form im Durchschnitt von ca. 40% (p-Wert < 0.005%; auch nach ad-hoc Korrektur signifikant:  $p = 3.5\%$ ). Die Fehlerverringeringung war im Zahlendiktat stärker ausgeprägt (Reduktion um ca. 70%) und statistisch hochsignifikant ( $p < 0.005\%$ ).

Diese drei frühen pädagogischen Studien beschreiben einheitlich eine Reduktion der Fehlerzahl um 40% bis 50% bei Einsatz einer stellenwertgerechten Sprechweise, tendenziell stärker ausgeprägt im Zahlendiktat als bei Kopfrechenaufgaben.

Die Ergebnisse dieser drei Studien wurden in neueren Untersuchungen grundsätzlich bestätigt, wo sich im Zahlendiktat bei Einsatz einer unverdrehten Sprechweise Absenkungen der Fehlerraten um 76% (Schmid 2023: 55 Kinder aus dem 2.

Schuljahr,  $p < 0,1\%$ ) und 67% (Hartwig 2025: 40 rechenschwache Grundschul Kinder,  $p < 0.5\%$ ) ergaben.

Die vorliegende Sekundäranalyse der drei frühen pädagogischen Studien erhärtet somit die plausible Hypothese, dass die Verwendung einer unverdrehten Zahlensprechweise im Grundschulmathematikunterricht zu deutlichen Vorteilen führt. Dies Ergebnis ist insbesondere wichtig, da zwei Arbeiten (Schellenberger 1953, Ammareller 2006) auch Kopfrechenaufgaben umfassten, also nicht nur Zahlendiktate, und zwei Arbeiten (Zehner 1955, Ammareller 2006) die Wirkung einer längeren Einübungszeit in die stellenwertgerechte Sprechweise beurteilen konnten.

## 7. Literaturverzeichnis

Ammareller MK (2006) Die nicht-invertierte Zahlensprechweise im arithmetischen Anfangsunterricht am Beispiel der Waldschule Bochum. Schriftliche Hausarbeit im Rahmen der Ersten Staatsprüfung für das Lehramt Primarstufe, Erziehungswissenschaftliche Fakultät der Universität zu Köln.

Ammareller MK (2008) Die nicht – invertierte Zahlensprechweise im arithmetischen Anfangsunterricht – ein Versuch. In: Gerritzen L (2008) Zwanzigeins - für die unverdrehte Zahlensprechweise, S. 65-71. Brockmeyer, Bochum. URL: <https://zwanzigeins.jetzt/infos/literatur>. Zugegriffen: 29.01.2025.

Armitage P, Berry G (1988) Statistical Methods In Medical Research. Oxford: Blackwell.

Cameron AC, Trivedi PK (2005) Microeconometrics. Methods and Applications. Cambridge: University Press.

Cameron AC, Trivedi PK (2010) Microeconometrics Using Stata. Texas: Stata Press.

Eiskirch S (2008) Chancengerechtigkeit im Mathematik-Unterricht der Waldschule. In: Gerritzen L (2008) Zwanzigeins - für die unverdrehte Zahlensprechweise, S. 62-

64. Bochum: Brockmeyer. URL: <https://zwanzigeins.jetzt/infos/literatur>. Zugegriffen: 29.01.2025.

Hartung J, Elpelt B, Klösener KH (2009) Statistik – Lehr- und Handbuch der angewandten Statistik, 15. Auflage. München: Oldenbourg.

Hartwig V (2025) Zwanzigeins vs. einundzwanzig: Profitieren rechenschwache Kinder in Bezug auf die Fehlerzahl von einer stellenwertgerechten Sprechweise? Eine empirische Untersuchung. Masterarbeit in Integrativer Lerntherapie an der Pädagogischen Hochschule Schwäbisch Gmünd, Zentrum für Wissenstransfer. <https://zwanzigeins.jetzt/aktivitaeten/projekte/unterrichtsversuche>. Zugegriffen: 29.01.2025.

Neter J, Wasserman W, Kutner MH (1985) Applied Linear Statistical Models. Homewood: Irwin.

Rothman KJ, Greenland S, Lash TL (2008) Modern Epidemiology, 2. Auflage. Philadelphia: Lippincott, Williams und Wilkins.

Schellenberger A (2008) Vierzig und Acht – Ein Pionier der Zahlensprechweise (Erinnerungen an meinen Vater). In: Gerritzen L (2008) Zwanzigeins - für die unverdrehte Zahlensprechweise, S. 38-47. Brockmeyer, Bochum. URL: <https://zwanzigeins.jetzt/infos/literatur>. Zugegriffen: 29.01.2025.

Schellenberger M (1953) Zahlwort und Schriftbild der Zahl. Leipzig: VEB Bibliographisches Institut.

Schmid S (2023) Zwanzigeins - Eine empirisch-quantitative Untersuchung zur Zahleninversion in der zweiten Schulstufe. Masterarbeit in Erziehungswissenschaft, Primarstufe. Kirchliche Pädagogische Hochschule Wien/Krems. <https://zwanzigeins.jetzt/aktivitaeten/projekte/unterrichtsversuche>. Zugegriffen: 29.01.2025.

StataCorp (2015) Stata - Release 14. Statistical Software. College Station, TX: StataCorp LP.

Zehner (1955) Das invertierte Zahlensprechen als pädagogisch-psychologisches Problem. Habilitationsschrift, Fakultät für Mathematik und Naturwissenschaften der Technischen Hochschule Dresden.

Zwanzigeins e.V. (2025a). Auch zwanzigeins und nicht nur einundzwanzig!  
<https://zwanzigeins.jetzt/>. Zugegriffen: 29.01.2025.

Zwanzigeins e.V. (2025b). Positionspapier.  
<https://zwanzigeins.jetzt/aktivitaeten/projekte/vorschlag-zahlensprechweise>.  
Zugegriffen: 29.01.2025.

Zwanzigeins e.V. (2025c). Zwanzigeins-App. <https://zwanzigeins.jetzt/app/index.html>.  
Zugegriffen: 29.01.2025.

Zwanzigeins e.V. (2025d). Unterrichtsversuche.  
<https://zwanzigeins.jetzt/aktivitaeten/projekte/unterrichtsversuche>.  
Zugegriffen: 29.01.2025.