



# Mathe macht stark/Lesen macht stark Evaluationsdesign

Benjamin Nagengast, Tobias Dörfler, Andreas Schulz,  
Sven Rieger

Auftaktveranstaltung, Hedwig-Dohm-Schule, Stuttgart  
03.07.2018



## Warum eine Evaluation des Modellversuchs?

- Qualitätssicherung von Reformen und Veränderungen im Bildungssystem
- Ziel: Evaluation von Modellversuchen = Entscheidungshilfe für Ausweitung und Implementierung
  - Werden die Ziele des Modellversuchs erreicht?
  - Unter welchen Bedingungen lassen sich die Maßnahmen besonders wirkungsvoll umsetzen?
  - Für welche Schülerinnen und Schüler sind die Maßnahmen besonders wirksam?



---

## Rahmenbedingungen der Evaluation von LMS/MMS

- Einbezug von Hauptschulen/Werkrealschulen, Gemeinschaftsschulen und Realschulen
- Integrative Förderung im Klassenverbund, Schulen bei Umsetzung mit Freiheiten
- Aufnahme aller angemeldeter Schulen in eines der Förderprogramme
- Beginn in Klassenstufe 5 und danach aufwachsend



## Fragestellungen der Evaluation

- Verbessern die beiden Förderprogramme LMS und MMS die Leistung der teilnehmenden Schülerinnen und Schüler?
  - (durchschnittlicher Effekt)
- Welche Schülerinnen und Schüler profitieren besonders?
  - (differentielle Effekte)
- Unter welchen Bedingungen sind die beiden Förderprogramme effektiv?
  - (Implementation der Programme)
- Zentrale Variablen:
  - Leistung, Motivation und Interesse



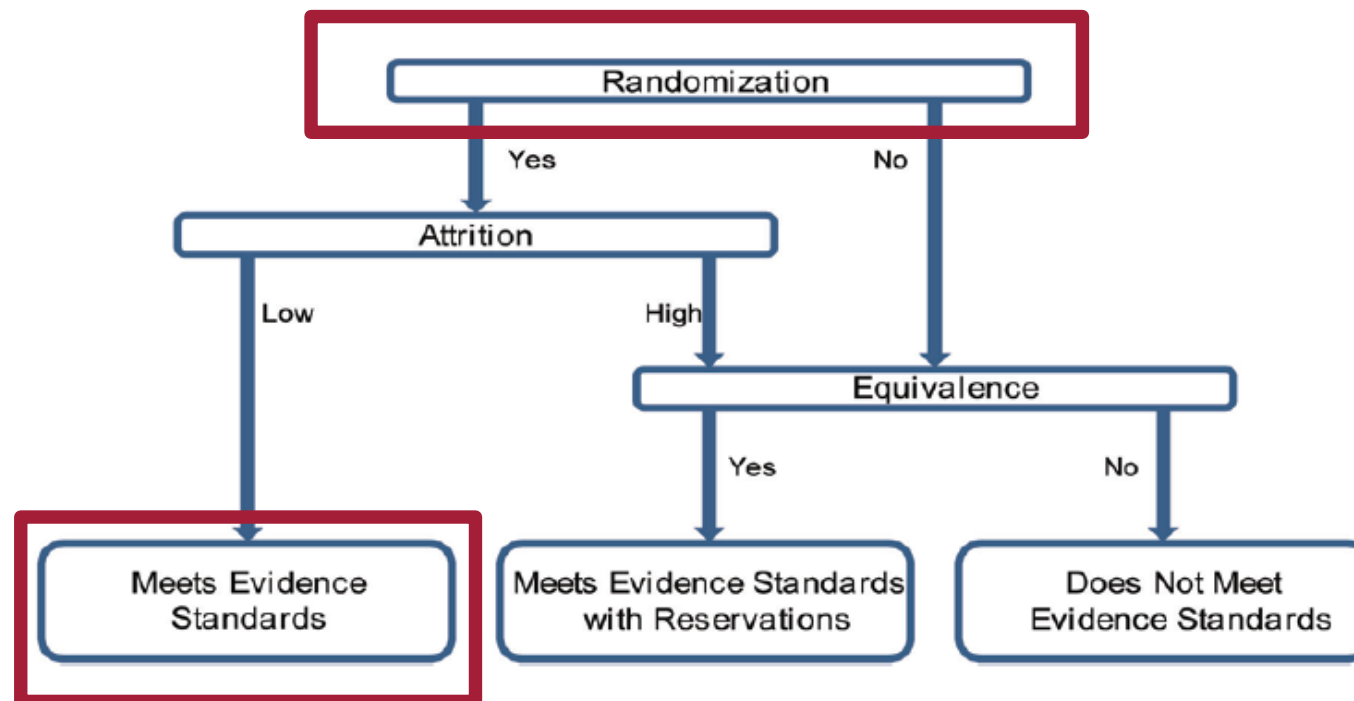
## Anforderungen an das Evaluationsdesign

- Unterschiede zwischen LMS- und MMS-Schulen sollen auf Teilnahme am jeweiligen Programm und nicht auf andere Faktoren zurückführbar sein (hohe interne Validität)
- Hinreichend große Stichprobe, die belastbare Aussagen ermöglicht (hohe statistische Validität)
- Erfassung der Implementierung der Programme als zentralen Wirkfaktor



## Wie sollte eine Evaluation aussehen?

- “What works clearinghouse“ → US Department of Education
- Nur Programme, die in randomisierten, kontrollierten Feldstudien getestet wurden können „Bestnoten“ für Evidenz erhalten





## Wie sollte eine Evaluation aussehen?

### Kontrollierte, randomisierte Studien:

- Kontrollierte Experimente:
  - Studiengruppe (mit Behandlung)
  - Vergleichsgruppe (ohne Behandlung)
- Randomisierte Experimente:
  - Teilnehmer der Untersuchung werden per Zufall Studiengruppe bzw. Vergleichsgruppe zugeteilt
  - Vorteil: Umwelteinflüsse & Teilnehmermerkmale (Alter, Herkunft, Motivation etc.) sind gleichmäßig auf die Gruppen verteilt



## Wie sollte eine Evaluation aussehen?

- Nachher Messung:



Veränderung? 





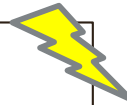


## Wie sollte eine Evaluation aussehen?

- Vorher-nachher Messung:



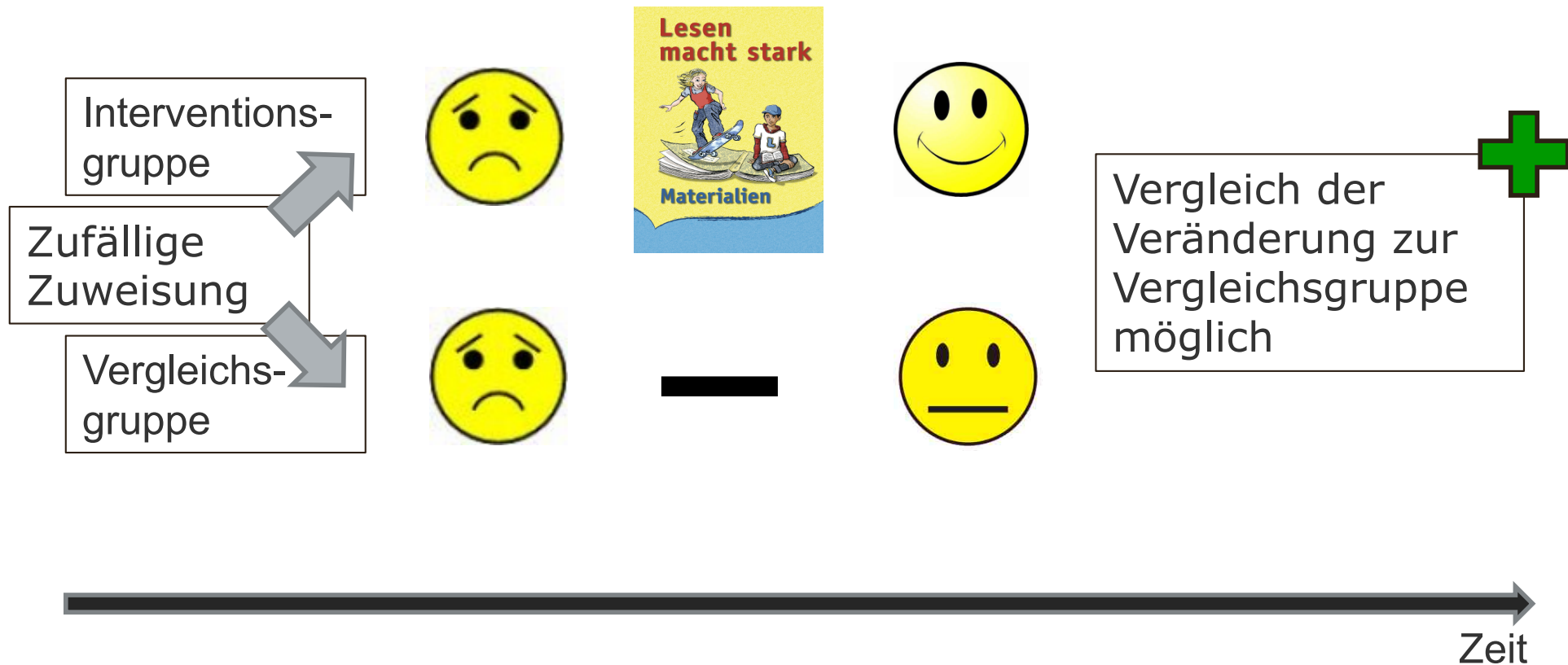
Worauf ist die  
Veränderung  
zurückzuführen?





## Wie sollte eine Evaluation aussehen?

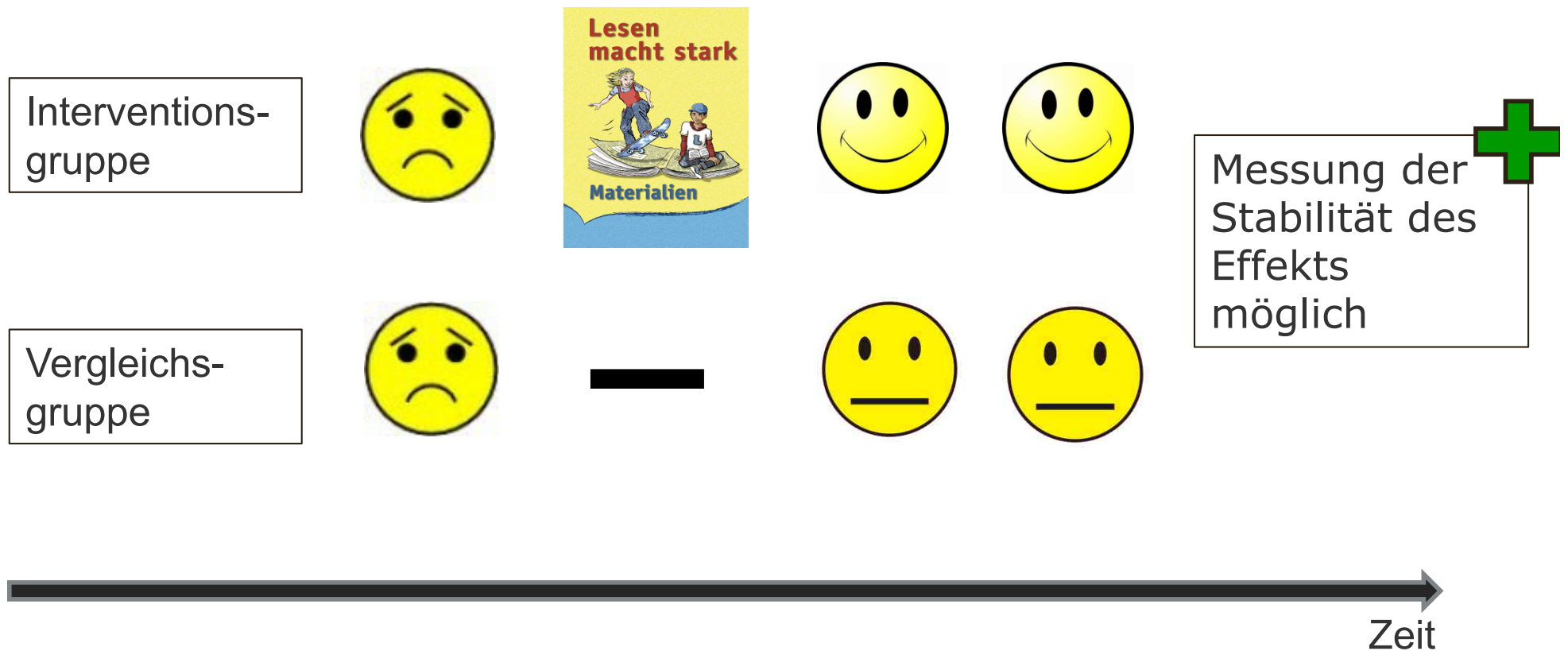
- Randomisiertes Vergleichsgruppendesign:





## Wie sollte eine Evaluation aussehen?

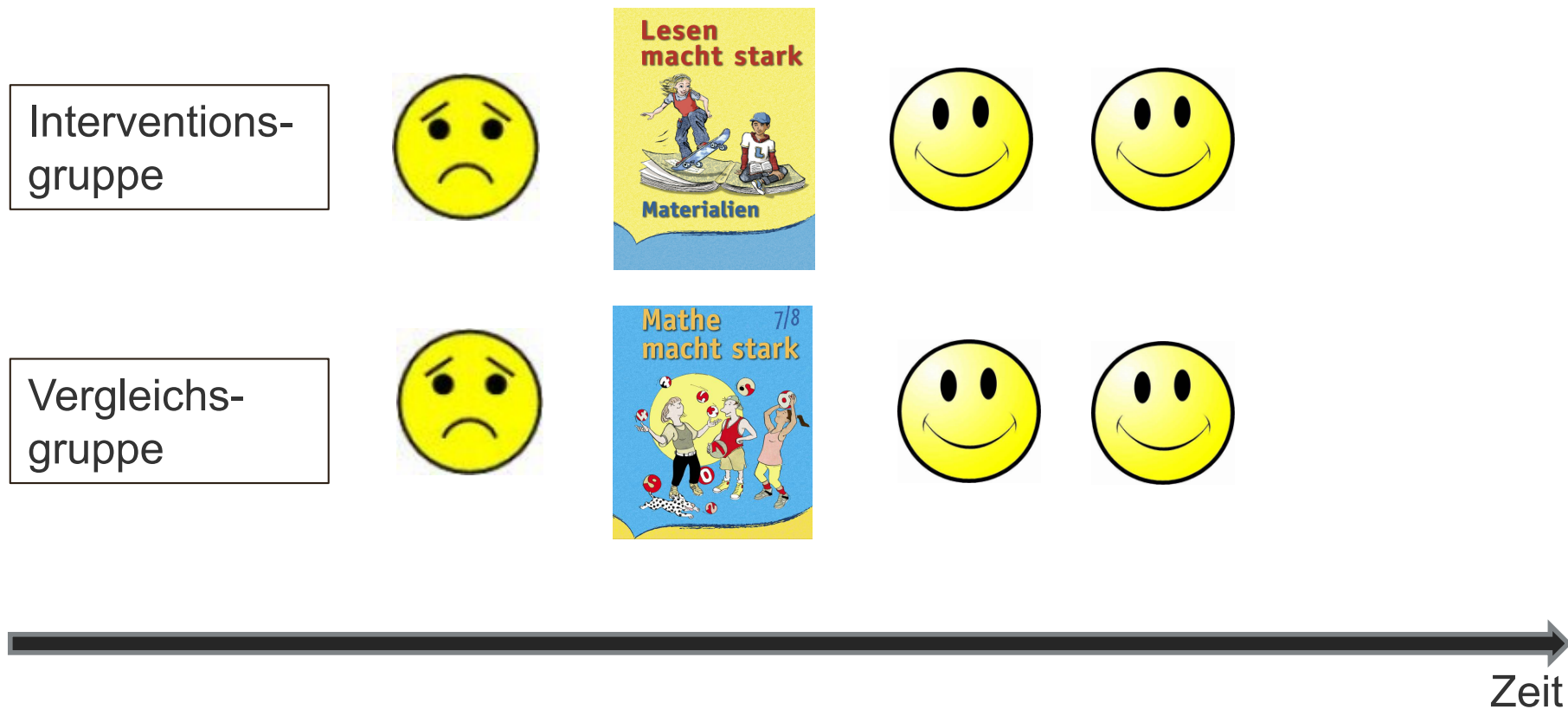
- Design mit Erfassung der Stabilität:





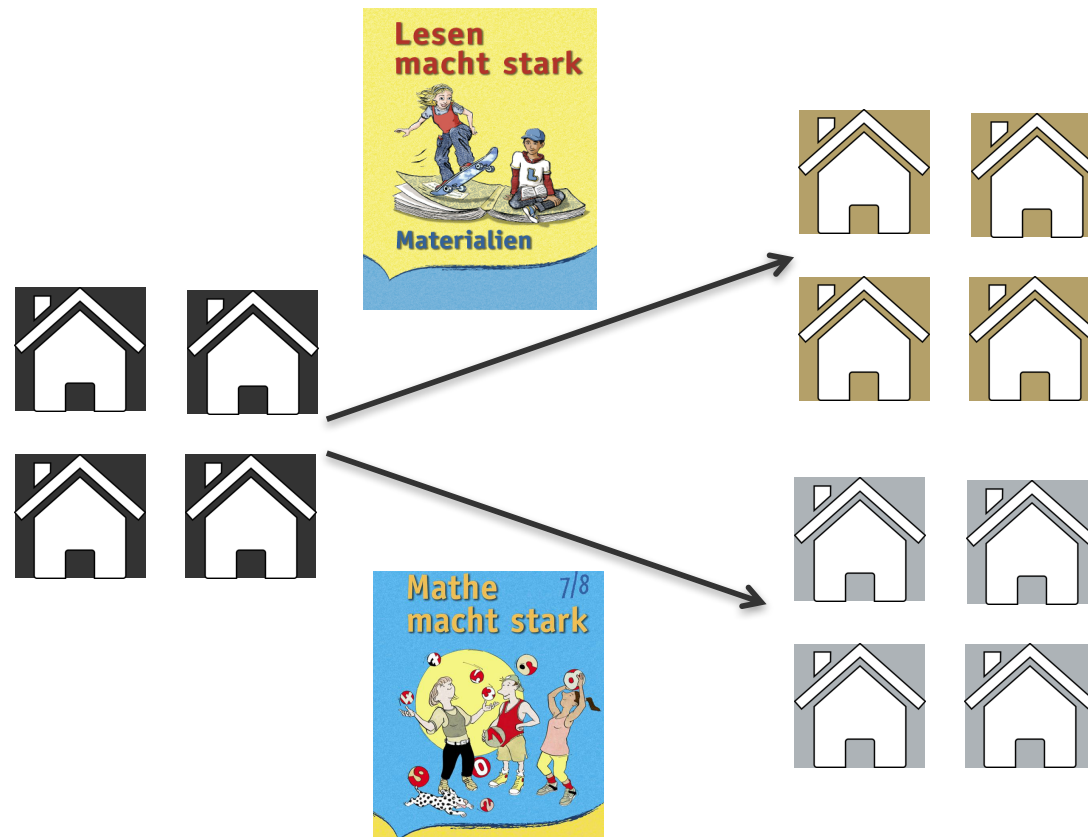
## Wie sollte eine Evaluation aussehen?

- Design mit Erfassung der Stabilität:





## Randomisierung von Schulen



- LMS und MMS dienen gegenseitig als Vergleichsgruppen

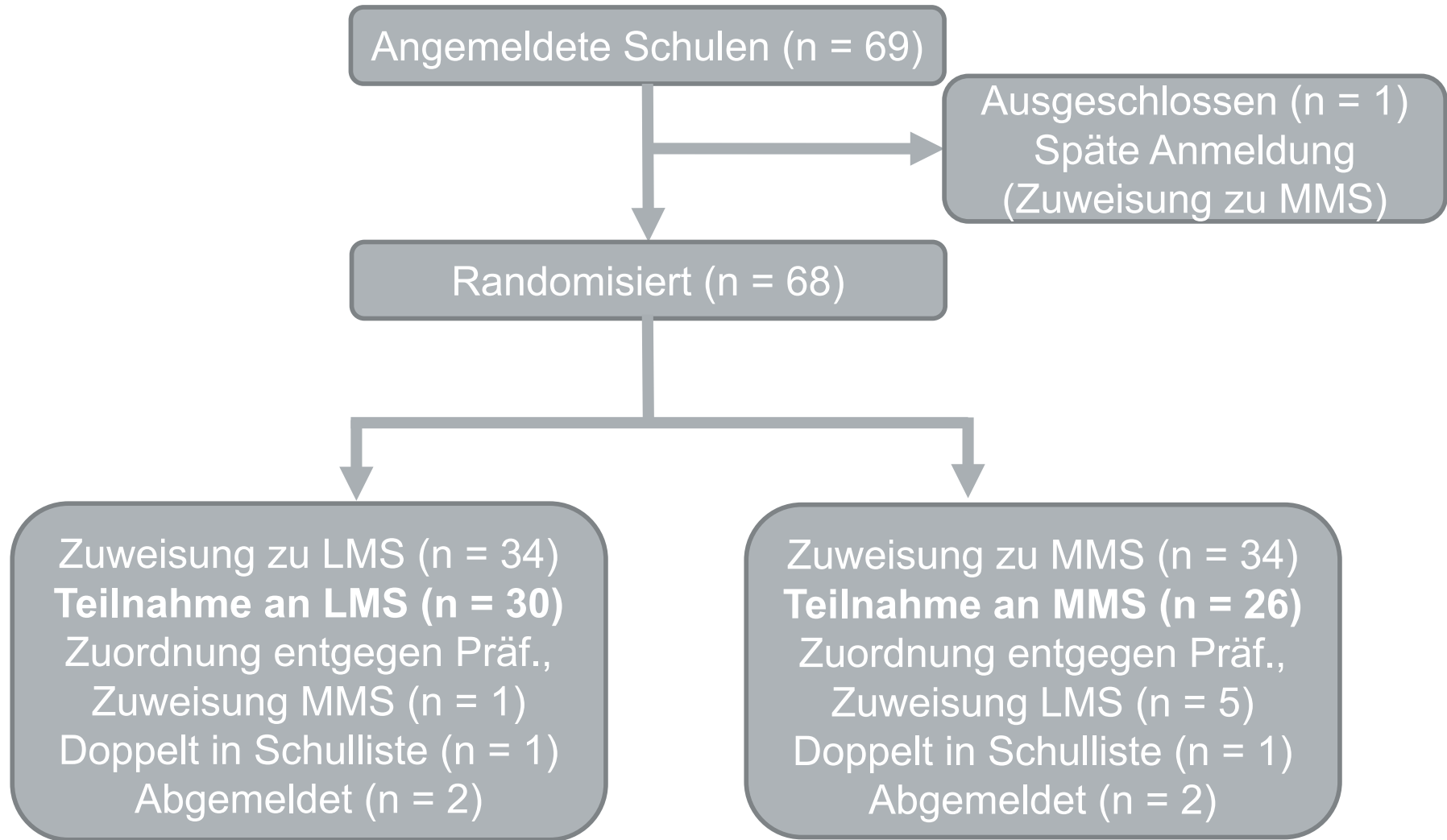


## Stichprobe

- Randomisierte Zuweisung innerhalb der Schulformen
- LMS
  - 5 Werkrealschulen
  - 14 Gemeinschaftsschulen
  - 11 Realschulen
  - 4 Schulverbände
- MMS
  - 6 Werkrealschulen
  - 14 Gemeinschaftsschulen
  - 10 Realschulen
  - 4 Schulverbände



## Stichprobe CONSORT-Diagramm





## Teststärke

- Wie wahrscheinlich ist es, den Effekt des Programms mit einer empirischen Studie „aufzudecken“, wenn dieser tatsächlich existiert?
- Ziel: Teststärke soll möglichst hoch sein (80%)
- Herausforderung:
  - Wie groß sind die Effekte, die wir erwarten können?
  - Wie gut können wir diese Effekte mit unserer Stichprobe aufdecken?





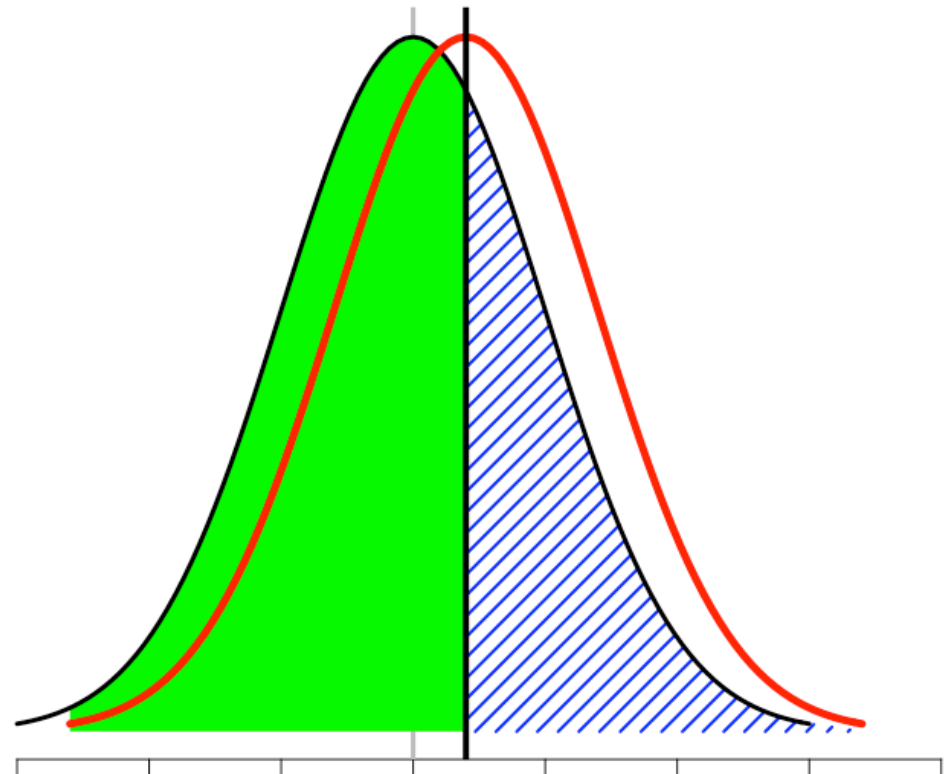
## Effektstärke

- Häufig gemessen als Cohens  $d$ 
  - Standardisierte Differenz der mittleren Leistung in der Interventions- und der Kontrollgruppe
- Approximative Effektstärken in Schleswig-Holstein:
  - LMS:  $d = 0.41$
  - MMS:  $d = 0.20$



## Effektstärke

- Effektstärke:  $d = 0.4$
- 66% der Kontrollgruppe unterhalb des Mittelwerts der Interventionsgruppe
- Zum Vergleich:
  - Leistungszuwachs durch ein Jahr Beschulung (z.B. Lüdtke et al., 2008)  
 $0.3 < d < 0.4$





## Teststärke

Berechnung der Teststärke für eine Effektgröße von  $d = 0.25$   
(PowerUpR; Bulus & Dong, 2018)

Annahmen:

$ICC = 0.3$ ,  $R2_{L1} = 0.5$ ,  $R2_{L2} = 0.5$ , 3 Kovariaten auf L2,  $\alpha = 0.05$

- Bei 56 Schulen
  - Teststärke  $(1 - \beta)$ : 0.64
  - Entdeckbare Effektstärke (MDES): 0.30



## Erhebungsdesign

	T1		T2	T3	T4
	September/ Oktober 2018	Februar/ März 2019	Juni/Juli 2019	Juni/Juli 2020	Juni/Juli 2021
Klasse 5	Prätest Lernstand 5	Implemen- tierung	Posttest 1		
Klasse 6				Posttest 2	
Klasse 7					Posttest 3

Evtl. weitere kleinere Erhebungen zur Erfassung der Implementierung



## Testverfahren

### Mathematik

- Lernstand 5 als Ausgangsmessung
  - Zahlverständnis, Operationsverständnis, (schriftliche Rechenverfahren)
- Entwicklung von Tests durch LS für Klassenstufen 6 und 7

### Lesen

- Lernstand 5 als Ausgangsmessung
  - Leseflüssigkeit, Leseverständnis
- Items aus ELFE II, verlinkt über Messzeitpunkte
- Lesetexte aus Evaluation in SH zur weiteren Verlinkung
- Lesestrategiewissen



---

## Schülerfragebogen

- Hintergrundvariablen
  - SES, Migrationsstatus, Deutschkenntnisse
- Allgemeine Schulzufriedenheit
- Motivation (jeweils für Lesen und Mathematik)
- Persönlichkeitsmerkmale
- Einschätzungen des Unterrichts aus Schülersicht



---

## Lehrerfragebogen

- Fachdidaktisches Wissen
- Motivationale Variablen (jeweils für Lesen und Mathematik)
- Überzeugungen zum Lernen im Fach



## Erhebungsumfänge (September/Oktober 2018)

- Lernstand 5
  - Zusätzliche Lesetests (45 Minuten)
  - Schülerfragebogen (45 Minuten)
  - Lehrerfragebogen
- 
- Erhebung durch Lehrkräfte in den Klassen vor Ort
  - Auswertung von Lernstand 5 durch Lehrkräfte über Portal des LS
  - Zusätzliche Tests/Fragebögen
    - Eingabe/Auswertung organisiert durch das KM





---

## Erfassung der Implementation

- Plan: Erfassung durch Fachberater Schulentwicklung anlässlich von Besuchen bei den Schulen vor Ort
- Beobachtung, Lehrerfragebogen und Schülerfragebogen



## Datenerhebung und Dateneingabe

- Teilnahme an der Evaluationsstudie ist für Schülerinnen und Schüler freiwillig
- Zustimmung der Eltern für Teilnahme an Evaluation und Weitergabe der Daten von Lernstand 5 erforderlich
- Verlinkung von verschiedenen Datenquellen (Lernstand 5, Tests, Fragebogen) erfolgt über Klassenlisten, die in Schule verbleiben
- Erstellung vor Sommerferien in Zusammenarbeit mit KM
- Konzept für Datenerhebung mit Datenschutzbeauftragtem des KM abgestimmt



## Ihre Rolle

- Fördern und unterstützen Sie Ihre Schülerinnen und Schüler mit den Materialien von LMS und MMS!
- Setzen Sie die Programme bestmöglich an Ihren Schulen um!
- Helfen Sie uns, belastbare Aussagen zur Effektivität der Programme und zu deren optimaler Implementierung zu machen!
  - Unterstützung der Evaluation
  - Ausfüllen der Lehrerfragebögen
  - Einverständniserklärungen
  - Organisation der Datenerhebung



# Kommentare und Fragen!

Prof. Dr. Benjamin Nagengast  
Universität Tübingen  
**Hector-Institut für Empirische Bildungsforschung**  
Europastraße 6, 72072 Tübingen  
Telefon: +49 7071 29-76089  
[benjamin.nagengast@uni-tuebingen.de](mailto:benjamin.nagengast@uni-tuebingen.de)