

Kerncurricula, Regelstandards, Kompetenztests:

Ab wann ist jemand eigentlich
"funktional alfabetisiert"?

Bericht von Hans Brügelmann über sein
Forschungssemester 2004/2005
vor dem Fachbereich 2 am 10.11.04

Schwerpunkte des Forschungssemesters

- Lehrbuch „Schule verstehen - Forschungsbefunde zu Kontroversen über Unterricht und Erziehung“
- Reanalyse der LOGIK-Studie (MPI München) zum „Prognoserisiko von Risikoprososen“ bei LRS
- Auswertung „Leseuntersuchung mit dem Stolperwörter-Lesetest“ (LUST-1) und Durchführung von Erweiterungsstudien (LUST-2)
- DFG-Antrag „Leseentwicklung von Grundschulkindern in verschiedenen Leistungsgruppen“

Erweiterung der LUST-Studien

Überprüfung der Befunde aus LUST-1 (2003):

- Längsschnitt statt Querschnitt
- Baden-Württemberg statt Nordrhein-Westfalen
- Validierung im Vergleich mit anderen Lese-Tests
- Validierung an der Leistung von Erwachsenen
- Differenzierung in 10 Untertests
- Übersetzung in MigrantInnen-Sprachen

Dank an ...

- den Fachbereichsrat FB2 und das Rektorat Uni Siegen
- das FuN-Kolleg Baden-Württemberg, die PH Schwäbisch Gmünd, vor allem Erika Brinkmann
- die KollegINnen in EW, besonders an die Arbeitsgruppe Primarstufe

Drei Foki des heutigen Vortrags

- Die empirische Frage:
Wie steht es eigentlich mit der „Lesekompetenz“ in Deutschland?
- Die politische Frage:
Hat das Schulsystem versagt - und müssen wir deshalb seine Regelungsmechanismen verändern?
- Die theoretische Gelenkstelle zwischen beiden:
An welchen Maßstäben messen wir „Lesekompetenz“, und wo setzen wir den Schwellenwert für „Lesenkönnen“?

Die zentrale Botschaft vorweg

- Wir können empirisch nicht begründen, welche Mindestleistungen SchülerInnen zu welchem Zeitpunkt erbringen müssen.
- Ein Systemwechsel von der „Input“- zur „Output“-Steuerung ist nicht notwendig und seine Nebenwirkungen sind vermutlich sogar kontraproduktiver als der erhoffte Ertrag.

(0) Vom „Input“ zum „Output“

- „Input“-Steuerung erfolgt über Lehrpläne, Lehrerausbildung und Prozesskontrolle durch Schulaufsicht.
- „Output-Steuerung“ definiert erwartete Leistungen und überprüft deren Erreichung mit Tests.

0.1 „Das ‚Input‘-System hat versagt, denn...“

- Deutschland hat bei PISA schlecht abgeschnitten.
- Deutschland hat ein Input-System.
- Erfolgreiche Länder haben Output-Systeme.
- Also ist das Input-System schuld an der PISA-„Katastrophe“.
- Wenn Deutschland auch ein Output-System einführt, werden die Schülerleistungen besser.

0.2 ... oder doch nicht?

Denn...

- Deutschland hat bei IGLU gut abgeschnitten.
- Die Grundschule und die Sekundarstufe arbeiten beide in einem Input-System.
- Also kann das Steuerungssystem nicht die Ursache für den PISA-„Misserfolg“ und zugleich für den „IGLU“-Erfolg sein.

0.3 Ist's eher der Zeitgeist?

- Kerncurricula → Kanon-Verpflichtung
- Regelstandards → Ziel-Vorgaben
- Kompetenztests → Erfolgskontrolle

→ „Das vermessene Kind“ (Jörg Ramseger)

→ „Der standardisierte Schüler“ (Kaspar
Spinner)

(1) Wie steht es mit der Lesekompetenz deutscher SchülerInnen tatsächlich?

Tab. 1

Internationale Vergleiche Rangplatz	IEA '91 16 Länder 9 Jahre	IGLU '02 35 Länder 10 Jahre	PISA '00 31 Länder 15 Jahre	IALS '93 8 Länder 16-65 Jahre
D'land	12	4-12	21	2
USA	2	4-12	15	7

In die Schule sind sie gekommen...

1934-1982	IALS (relativ konstant):	Deutschland +
1988	IEA	Deutschland -
1991	PISA	Deutschland -
1998	IGLU	Deutschland +

Beliebigkeit der Urteile

Und nicht nur die internationalen
Rangplätze -
auch die Einschätzungen der
Anteile von „Risikogruppen“
in der Gesamtpopulation
schwanken erheblich →

Ab wann ist jemand ein „Risiko“-Schüler?

Studie	Alter	Jahr	Anteil „Risiko“
IEA	14	1991	2 %
PISA	15	2000	25 %
IGLU	9	2003	10 %
OECD	16+	1993	14 %
ANALFA	16+	1980	5 %

Das Forschungsproblem

Damit stellt sich die Frage
nach der Begründung

- von Maßstäben und
- von Kriterien

für die Bestimmung von
Schwellenwerten.

1.1 Zur alltagsökologischen Validität von Schwellenwerten

Die International Adult Literacy Study der OECD (1995) war ein erster Versuch, Anforderungen an die Lesekompetenz im Alltag durch Erhebung von objektiven Leistungen und subjektiven Einschätzungen der eigenen Kompetenz genauer zu bestimmen.

Selbsteinschätzung der Kompetenz auf verschiedenen Leistungsstufen

Tab. 7 (nach: OECD 1995)

Selbsteinschätzung Alltag (Deutschland)	Stufe 1	Stufe 4/5
gut+	81 %	99 %
mittelmäßig	12 %	-
schlecht	5 %	-

Selbst wenn man soziale Erwünschtheit unterstellt, überraschen die 81% „gut+“ auf Stufe 1

Schriftsprachliche Aktivitäten auf verschiedenen Leistungsstufen

Tab. 8 (nach: OECD 1995)

Deutschland	Stufe 1	Stufe 4/5
Zeitung lesen	90 %	97 %
Bücher lesen	35 %	64 %
Briefe schreiben	13 %	27 %
Bibliothek besuchen	12 %	36 %

**Kompetenz und Häufigkeit korrelieren –
aber die Unterschiede zeigen sich vermutlich
eher in der Qualität als in der Quantität**

Niedrige Kompetenz nach Leistungsgruppen & Textsorten

Tab. 11 (nach: OECD 1995)

LeserInnen (Deutschland) Anteil Stufe 1	narrative Texte	Gebrauchs- Texte
deutschsprachig	12 %	8 %
MigrantInnen	38 %	23 %

**Lesekompetenz variiert je nach Textsorte –
und: Lesen lernt man auch außerhalb der
Schule → MigrantInnen Gebrauchstexte**

Überlappung der Lesekompetenz in verschiedenen Gruppen

Stolperwörter-Sätze: Richtige Sätze/ Min.					
Gruppe	Mitte 2. Kl.	Mitte 4. Kl.	Berufs- schule	Hand- werk	Leh- rer
N =	6.654	6.415	252	166	181
aM	4.1	8.1	14.9	11.2	18.7
SD	2.1	2.7	3.1	3.4	3.7
Min-Max	0-11	1-16	6-25	3-20	9-29

1.2 Zur lernbiografischen Validität von Schwellenwerten

Bisher gibt es kaum empirisch abgesicherte Daten zu „Mindestleistungen“ auf einer bestimmten Jahrgangsstufe als Voraussetzung für späteren Schulerfolg.

In zwei Studien haben wir Möglichkeiten überprüft, die Leseleistung von SchülerInnen höherer Klassen durch die Bestimmung von „Voraussetzungen“ auf früheren Jahrgangsstufen vorherzusagen.

Ist das Schulversagen „der unteren 15%“ vorprogrammiert?

- Prognoseversuche über eine Klassifikation von Risikokindern am Schulanfang:
Unsere Reanalyse der LOGIK-Studie von Weinert/ Schneider (1999)
- Studien zum Lernzuwachs über die Grundschulzeit hinweg:
Unsere Jahrgangs-Querschnitte und Individual-Längsschnitte im Projekt LUST

Vorhersage der Rechtschreibleistung im LOGIK-Projekt

Prädiktor vor Schul- anfang	Recht- schreibung < 15 %	Prognose Treffer- quote
BISC < 15%	8 Jahre	59 %
	10 Jahre	37 %
	17 Jahre	24 %

- **Lernverläufe sind durch individuelle „Voraussetzungen“ nicht vorhersagbar.**
- **Selbst bei erhöhtem Risiko (24% statt 15%) ist die jeweilige Erfolgswahrscheinlichkeit deutlich höher (76% zu 24%)**

Stabilität niedriger Positionen = Misserfolg?

Die Befunde aus LUST-1 und LUST-2 zeigen:

- Die Differenz zwischen den oberen 10% und den unteren 10% beträgt auf allen Jahrgangsstufen mehrere Schuljahre.
- Die Rangplätze bleiben im Längsschnitt (bisher über 6-12 Monate hinweg) relativ stabil.
- Die unteren Leistungsgruppen machen vergleichbare Fortschritte wie die oberen:
→ „Karawanen-Effekt“

Zwischenbilanz I

- Objektive Leistung und subjektive Zufriedenheit mit der eigenen Kompetenz können erheblich divergieren.
- Wir haben keine verlässliche Empirie zur Begründung von Schwellenwerten als „Voraussetzung“ für zukünftigen Schulerfolg.
- Mindestwerte beruflich erfolgreicher Erwachsener weichen beträchtlich von den Durchschnittswerten ihrer Gruppen ab.

Folgerungen I

„Bildungs“standards, die von allen SchülerInnen die gleiche Leistung zum gleichen Zeitpunkt fordern,

... lassen sich als Voraussetzung für erfolgreiche Teilnahme am Unterricht bisher nicht begründen;

... machen als Lernziel angesichts der unterschiedlichen Voraussetzungen wenig Sinn;

... sind als Maßstab für den tatsächlichen Lernerfolg (= individuellen Fortschritt) nicht brauchbar.

(2) Vom Input zum Output

- **Regelstandards → Ziel-Vorgaben**
- **Kompetenztests → Erfolgskontrolle**

(2.1) Kerncurriculum - Einwände...

= Input, aber mit Output-Erwartungen.

Dagegen:

- Verschiedene können Dasselbe an unterschiedlichen Inhalten lernen.
- Verschiedene lernen Verschiedenes an denselben Inhalten.
- Verschiedene sollten in der Schule Unterschiedliches lernen dürfen.
- Im Austausch können alle auch Anderes kennen lernen.

➔ Begegnung statt Belehrung oder gar Bekehrung

Durch Latein lernt man denken?

Problem des Transfers

Empirie (schon Thorndike 1923/24) zeigt, dass...

- nicht Denken an sich (z. B. Logik)
- sondern inhaltsbezogen (z. B. Wortschatz)

also in hohem Maße bereichsspezifisch gelernt wird

(2.2) Regelstandards

Unklarheit der Begründung ...

- als Verpflichtung der Schule -
oder als Bringschuld der SchülerInnen?
- über einen normativen Anspruch -
oder über empirische Belege?

2.2.1 Regelstandards - Einwände...

- Definition durch empirische Verteilung?
Es gibt „immer untere 15%“!
- Prognostische Validität nicht belegt:
 - keine lernbiografische Validität: P-S1
 - keine ökologische Validität: Alltag
- Sinn eines gleichen Anspruchs an alle...,
wenn →

2.2.2 Lernerfolg = Lernzuwachs

- wenn bereits am Schulanfang „Voraussetzungen“ um 3-4 Entwicklungsjahre streuen...
 - wenn weiter alle Schülergruppen vergleichbare Fortschritte machen („Karawanen-Effekt“) -
- Kann man dann Lern“erfolg“ am Erreichen gleicher Niveaus zu demselben Zeitpunkt festmachen?

(3) Kompetenztests

- **Deskriptiv hilfreich zur Einordnung von Leistung
- aber normativ problematisch**
- **Unklarheit der Funktion im Evaluationssystem**

3.1 Drei Funktionen als Anspruch

- Bestandsaufnahme des Systems
 - Kontrolle von LehrerInnen und Schulen
 - Diagnose von Leistungsständen und Lernschwierigkeiten bei SchülerInnen
- Überforderung des Instruments

3.1.1 System-Monitoring - Einwände...

- Landesweite Kompetenztests sind eine sinnvolle Ergänzung des Evaluationssystems,

aber es würden reichen:

- Wiederholungen alle 3-5 Jahre
- Stichproben statt Vollerhebungen
- Zusätzlich freiwillige Teilnahme für Interessierte.

3.1.2 Lehrer-Kontrolle - Einwände

- Teaching to the test: USA, England
 - Verarmung des Curriculum
 - Scheinanstieg im Testbereich
- Zunehmende Aussonderung der Schwachen
- Umwegstrategien: Vorbereitung, Glukose
- Täuschung durch Test(ver)fälschung (→ VERA)

3.1.3 Schüler-Diagnose - Einwände...

Probleme:

- punktuell statt kumulativ
- Zu kleiner Leistungsausschnitt
- Messfehler größer als bei kontinuierlicher Begleitung (und als bei Gruppenwerten)

Zwischenbilanz II

- Das System der Standards und landesweiten Tests wird mit zu vielen Erwartungen überfordert.
- Durch die Konzentration der Ressourcen auf der zentralen Ebene werden unterrichtsnahe Aktivitäten der Evaluation vernachlässigt.
- Notwendig ist die Entwicklung eines nach Ebenen und Funktionen differenzierten Beobachtungs- und Rechenschaftssystems.

Vom Produkt zum Prozess

Das Fatale an der aktuellen Diskussion über Leistungsstandards ist, dass sie sich auf die Definition und Messung des „Outputs“ beschränkt und keine Kriterien für die Qualität der Prozesse benennt oder diese überhaupt evaluiert.

Damit wird Oberflächenanpassung gefördert und nicht die viel beschworene Unterrichts“kultur“ entwickelt. Diese aber entscheidet darüber, was die gemessenen Leistungen tatsächlich wert sind.

Schluss-Thesen

Die aktuellen Bemühungen um eine Reform des Bildungssystems „nach PISA“ sind auf allen Ebenen zu simpel und zu mechanistisch gedacht:

- im Verständnis des Lernens von SchülerInnen als gleichschrittiger Kumulation von Kompetenzen;
- in der Forderung an LehrerInnen, SchülerInnen lernen zu „machen“ und
- im Glauben an eine „output“-orientierte Steuerbarkeit des Bildungssystems von oben.