

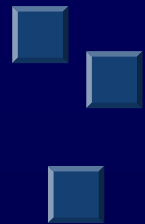


# Internationale Erfahrungen mit Output-Standards

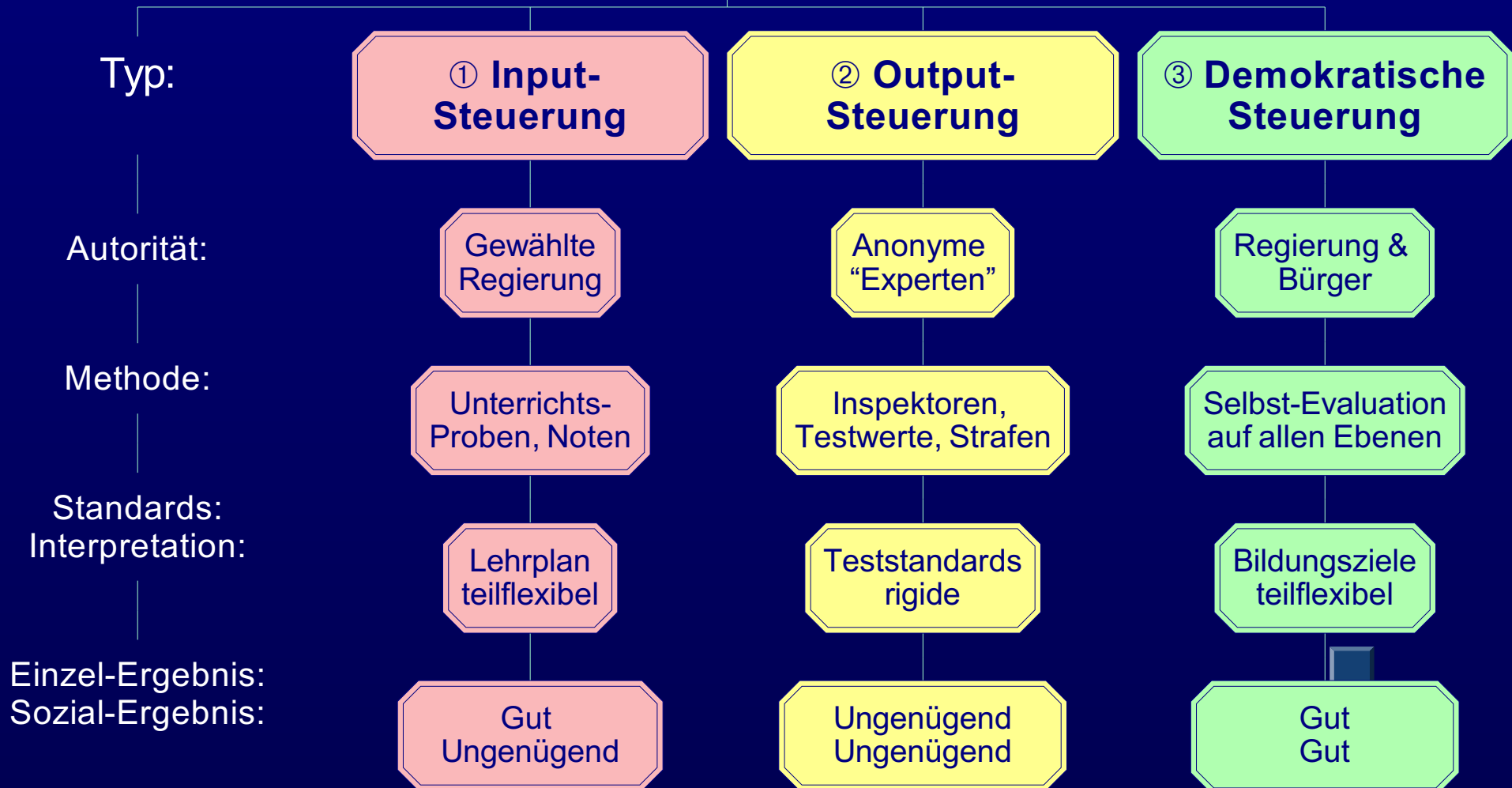
Prof. Dr. Georg Lind

Fachbereich Psychologie  
Universität Konstanz

Georg.Lind@uni-konstanz.de  
<http://www.uni-konstanz.de/ag-moral/>



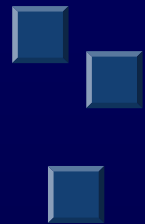
# Typen der Qualitätsentwicklung





# Beispiel für technokratische Qualitätsentwicklung: USA, I

- Die Theorie:
  - ▶ Output-Standards sollen die Lernleistungen verbessern, indem sie klar definieren, was unterrichtet werden soll, welche Verhaltensweisen von Lehrern und Schülern erwartet werden, und welche Sanktionen bei Versagen drohen.






# Beispiel für technokratische Qualitätsentwicklung: USA, II

## ■ Die Praxis

- ▶ 1965 - Die US-Bundesregierung unter Johnson gibt mit dem 'Head Start'-Programm erstmals direkt Input-Standards für Schulen vor.
- ▶ 1983 - Wegen angeblichem Verlust der Wirtschaftskraft ('A Nation at Risk') gibt die Reagan-Regierung Output-Standards für alle Schulen vor.
- ▶ 1991- Regierung Georg H. Bush: *National Education Goals 2000*
- ▶ 1994 - Regierung Clinton: *Goals 2000: Educate America Act*
- ▶ 2002 - Regierung Georg W. Bush: No-Child-Left-Behind Gesetz. Schulen müssen jedes Jahr höhere Testwerte vorweisen.

## ■ Das Ergebnis

- ▶ Kein nennenswerter Leistungsanstieg, Verengung des Lehrplans, April 2006: 30% Schulabbrecher (viele davon 'push-outs')
  - ▶ Vorhersage: 2014 werden 95% aller Schulen zu 'Versagern' gestempelt sein.
- 



# ... und jetzt auch in Deutschland

## Bildungsstandards

- “... die KMK [hat] Maßnahmen [...] beschlossen, die vor allem darauf abzielen, die *Unterrichtsqualität* zu verbessern und *nachhaltiges Lernen* zu fördern.
- Eine große Rolle spielen dabei die Bildungsstandards [...].
- [...] wie diese Ziele erreicht werden sollen, [wird] den Schulen [gemeint sind ‘Schulleiter’] überlassen.
- [...] in bestimmten Abständen werden die Arbeit und die Ergebnisse von Schulen evaluiert.”

Josef Erhard, Ministerialdirektor, Vorsitzender der Amtschefskommission  
„Qualitätssicherung“ in der Kultusministerkonferenz (KMK). [meine Einfügung, GL]

Quelle: [http://www.forumbildung.de/templates/imfokus\\_inhalt.php?artid=352](http://www.forumbildung.de/templates/imfokus_inhalt.php?artid=352)

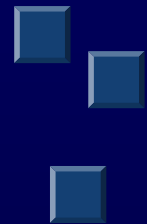




# Bildungsstandards

## aus technokratischer Sicht

- „Bildungsstandards [benennen] zentrale Ziele und Konzepte eines Fachs sowie
- die fundamentalen Kompetenzen, die Schülerinnen und Schüler
- bis zu einer bestimmten Jahrgangsstufe
- an zentralen Inhalten und Kernbereichen eines Fachs erworben haben sollen.
- Sie lenken damit die Aufmerksamkeit und die Praxis im Unterricht auf kumulatives Lernen, d.h. auf langfristig aufgebaute Lernergebnisse.“
  - Klieme, E. /Steinert, B.: Einführung der KMK-Bildungsstandards. MNU 57/3, 132-137 (meine Absatzgliederung, GL)



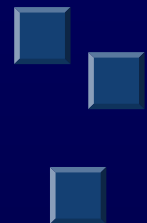


# Bildungsstandards = Teststandards

“Aufgabe des [Instituts für Qualitätsentwicklung im Bildungsbereich] IQB ist es, „...die Standards mit Leben [!] zu füllen, indem wir dafür eigene *Testaufgaben* entwickeln und präzise definieren, was wir von den Schülerinnen und Schülern erwarten.”


Prof. Köller, Direktor des IQB

Meine Ergänzungen in [ ]





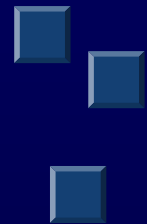
# Argumente für eine Output-orientierte Qualitätsentwicklung

- ① Deutsche Schulen seien nicht leistungsfähig
  - ② Ursache dafür sei die bisherig “Input”-Steuerung
  - ③ Bildungsziele ließen sich durch Standards ersetzen und durch Testaufgaben umfassend und valide feststellen
  - ④ Lernen verläuft linear oder ‘kumulativ’
  - ⑤ Daher könnten die *Schwierigkeit* und *Gleichwertigkeit* von Testaufgaben sowie die *Kompetenzstufen* allein durch *statistische* Kennwerte definiert werden
  - ⑥ Durch Bildungsstandards, Tests und Sanktionen könnten Lernleistungen verbessert und soziale Ungleichheiten verringert werden.
- 



Die Realität

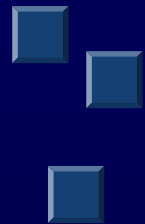
**Die technokratische  
Qualitätsentwicklung mittels  
Output-Standards hat sich bereits  
jetzt als Irrweg erwiesen**





# ① Deutsche Schulen sind besser als Tests sie machen

- PISA lässt aufgrund seiner Anlage kaum Aussagen über die Qualität eines Schulsystems zu
- Andere Vergleichsstudien zeigen deutsche Schüler zum Teil in der internationalen Spitzengruppe, zum Beispiel
  - in Bezug zu Mathe- und Lese-Fähigkeiten (z.B. IGLU)
  - in Bezug zu Geographiekennntnisse (National Geographic-Studie)
  - in Bezug auf moralisch-demokratische Kernkompetenzen (Lind, 2002)
- Die hohe Qualität deutscher Ausbildung zeigt sich auch in anderen Kennzahlen (hohe Außenhandelsüberschuss, hohe Produktivität, geringe Jugendkriminalität etc.)

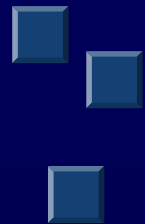




## ② Die Input-Steuerung hat sich weitgehend bewährt



- Die Inputsteuerung erlaubt eine weitgehend gleichmäßige Versorgung aller Kinder mit schulischer Bildung
- Die Input-gesteuerten Grundschulen können trotz einer materiellen Unterausstattung international gut mithalten
- Lehrpläne (Input) lassen dem Lehrer heute viel Spielraum
- Qualitätsmanagement nach ISO 9000 sind reine Input-Standards





### ③ Bildungsziele lassen sich nicht durch Standards ersetzen ...



- Bildungsziele umfassen viel mehr Kompetenz-Arten (behalten, verstehen, anwenden, verantworten, etc.) in viel mehr Wissens-und Lebensbereichen als in den Standards formuliert und durch Tests gemessen wird.
- Demokratische Bildungsziele betreffen nicht nur Erwartungen der Gesellschaft an die Fähigkeiten des Einzelnen, sondern auch Erwartungen des Einzelnen an sich selbst und an die Gesellschaft.

'Standards von unten' (Brügelmann, 2004).






# Output-Standards: Zwischen Hybris und Bescheidenheit


- Mit billigen Output-Tests wird nur ein sehr geringer Teil der für das Leben wichtigen Kompetenzgebiete gemessen ...
  - ▶ Lesen, Rechnen, Naturkunde
  - ▶ Es fehlen Geschichte, Politik, Geographie, Musik, Kunst, Kommunikation, Management, Verantwortung, Führung ...
- und auch nur jene Aspekte, die (ökonomisch) messbar sind
  - ▶ reines Begriffswissen (Fakten etc.) und einfache reproduktive Fertigkeiten (Rechtschreibung, Ausdruck, Rechenaufgaben)
- Kaum oder nicht gemessen werden höhere *creative* und *produktive* Fähigkeiten, sowie der Fähigkeit, zu schreiben, Wissen zu *verstehen*, *anzuwenden* und zu *verantworten*. 
- Nicht messbar sind Kompetenzen, die sich erst später im Leben zeigen können (zB Verantwortung im Beruf oder als Eltern). 



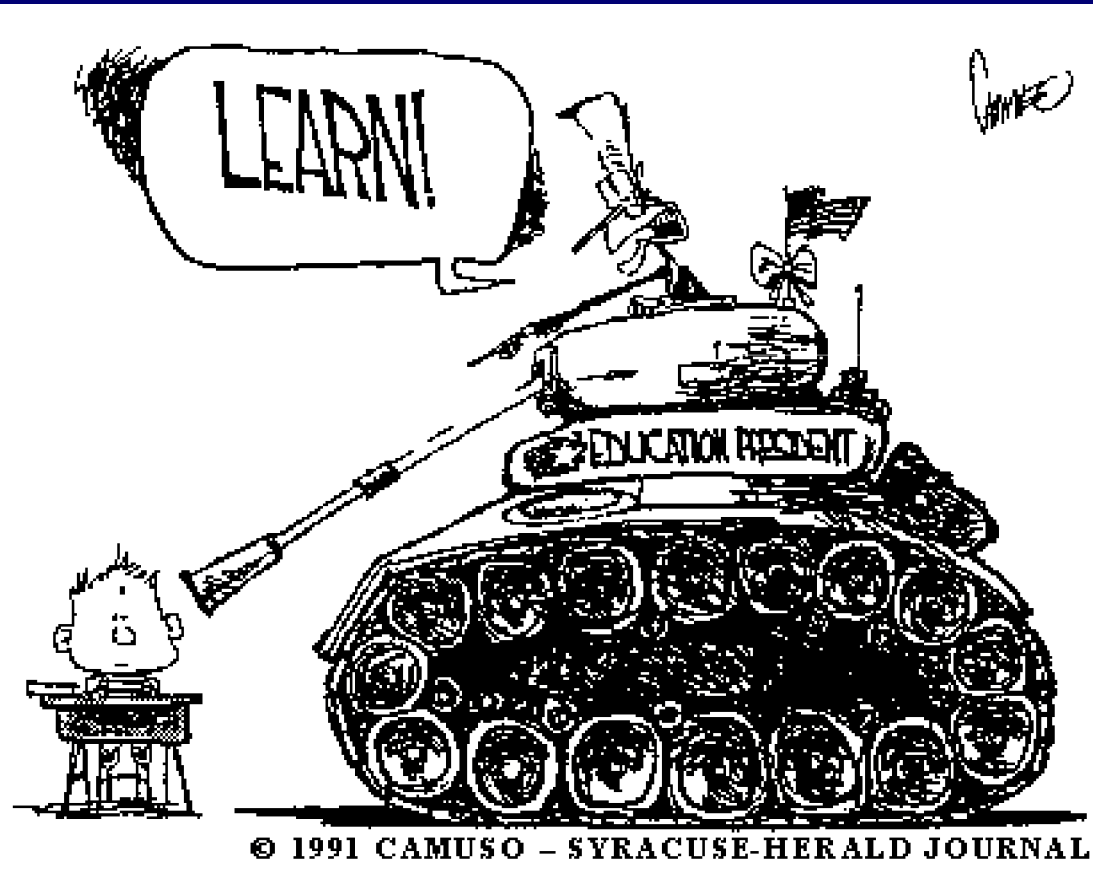
# Die Welt der Output-Standards



Ergebnisse der Erhebung der Organisation *Mid-Continent Research for Education and Learning*, 2004 (<http://www.mcrl.com>)

- In Dokumenten von diversen Schulgremien auf kommunaler, Lands- und Bundesebene in den USA wurden 4.100 verschiedene Standards ermittelt
    - ▶ Davon betrafen
      - 3174 Inhalte (77,4%) betrafen rein deklaratives Wissen, also reines Bücherwissen;
      - 799 Inhalte (19,5%) handelten von prozeduralem Wissen, als von “Wissen-Wie”
      - Nur 127 Inhalte (3%) betrafen kontextuelles Wissen, also anwendungsbezogene Kompetenzen
    - ▶ Keine Inhalte betrafen Handlungs- und Verantwortungswissen, also Wissen, das unter sozialem und emotionalem Druck verfügbar ist
- 

## ⑤ Die Lern-”Theorie” hinter den Output-Standards



“Lerne!”

Der Bildungs-  
Präsident

# Kritik der Output- Standards aus Sicht der experimentellen Lehr-Lernforschung

- Keine Fundierung durch kohärente, empirisch bewährte psychologisch-pädagogische Theorie
- Meist rein empiristische “Theorien”
- Keine Validierungsstudien an prognostischen Kriterien
- Pseudo-Praxis-Bezug der Testaufgaben
- Extreme Textlastigkeit der Testaufgaben in Mathematik und Naturwissenschaft ( $r = 0,84$  bzw.  $0,87$ )
- Heimliche Lerntheorie mit falschen Annahmen





# Testaufgaben



## Ein Beispiel für heimliche Lernpsychologie

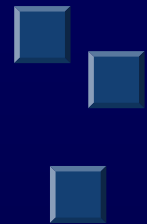
Zum Beispiel: Die Schüler sollen ausrechnen, ob es eine Zeitersparnis bringt, wenn ich auf einer Hauptstraße von A nach B mit 50 km/h fahre oder einen abkürzenden Schleichweg mit 30 km/h.

Das man so etwas ausrechnen können sollte, mag ja sein, aber die Schüler/innen sollten doch wohl etwas anderes lernen: nicht ausgerechnet mal wieder nach dem Faktor Zeit gucken (Eile und Stress!), sondern ob der Verkehr glatt läuft (Rücksichtnahme), ob der Schleichweg durch ein Wohngebiet führt (Gefährdung von spielenden Kindern!), welche Fahrweise am wenigsten Treibstoff verbraucht (Umweltbewußtsein).

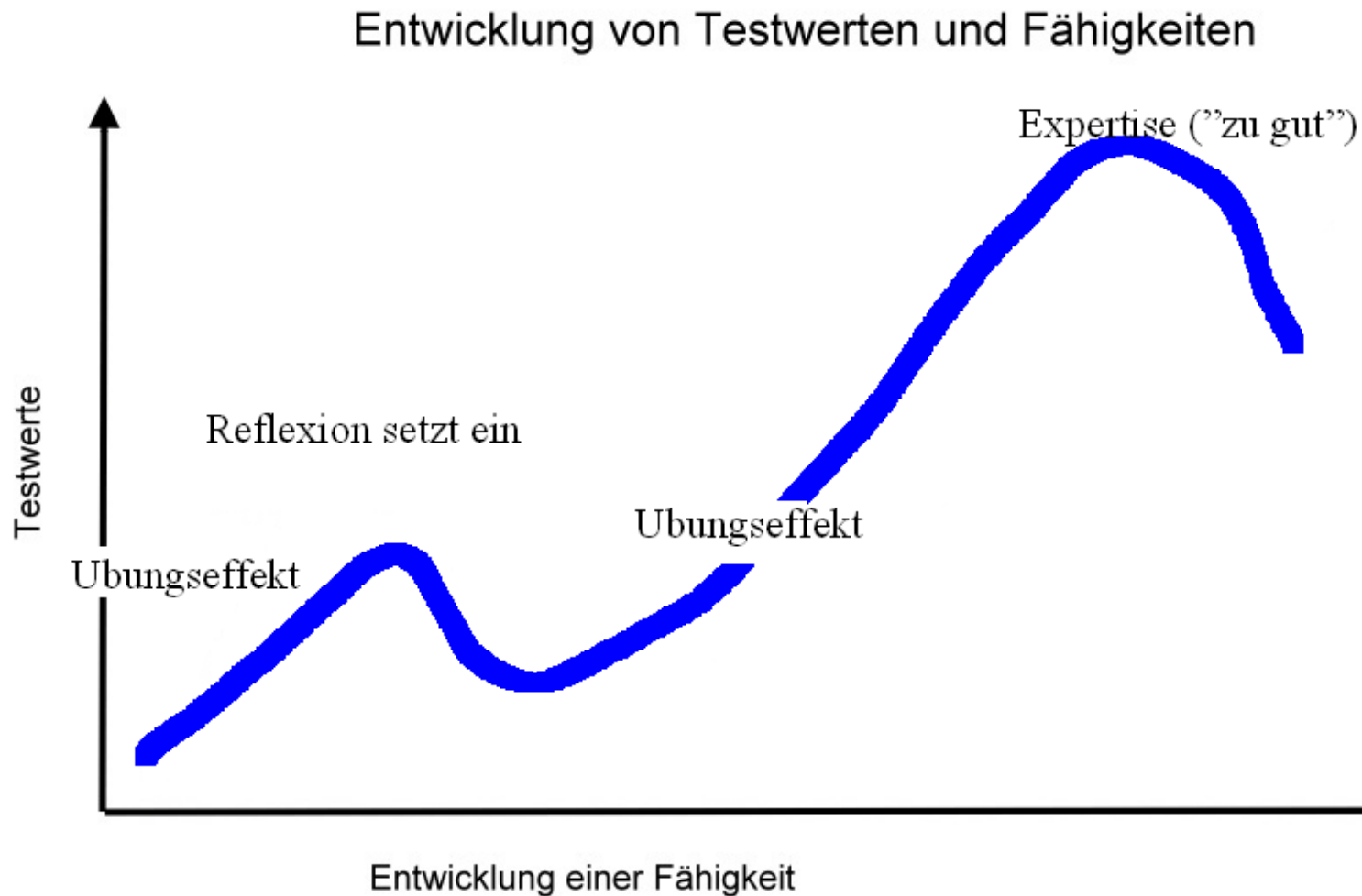


# Die Sicht der modernen Lernpsychologie

- Lernen ist oft nicht *kumulativ und linear*:
  - ▶ Die Lösung 'schwieriger' Aufgaben setzt nicht immer die Lösung 'leichter' Aufgaben im Sinner der Testtheorie voraus.
  - ▶ 'Kompetenzstufen' können übersprungen oder durch andere Lernwege ersetzt werden, vor allem in breiten Lernfeldern
- Wie 'einfach' oder 'komplex' Aufgaben sind, liegt im 'Auge des Betrachters'
  - ▶ Kinder können angeblich 'komplexere' Fähigkeiten beherrschen, bevor es einfache können. Beispiel: Lesen durch schreiben
  - ▶ Wer aus Sicht der Tests zu viel weiß, bekommt weniger Punkte.



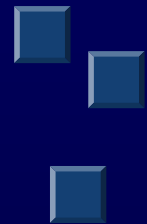
# Es gibt keinen linearen Zusammenhang zwischen Testleistung und Fähigkeit





# Wie bei Tests die Realität der Theorie angepasst wird

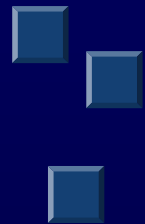
- Die 'Kumulativität' und 'Linearität' des Lernens werden künstlich hergestellt, indem alle Aufgaben eliminiert werden, die dieser Annahme widersprechen ('Testanalyse', 'Verbesserung der Standards').
- Die 'Schwierigkeit' und 'Gleichbedeutung' von Aufgaben wird nicht durch fundierte psychologische oder didaktische Theorie bestimmt, sondern bloß durch den Vergleich von Lösungshäufigkeiten.





## ⑥ Bildungsstandards verhindern besseren Unterricht und nachhaltiges Lernen

- Der Unterricht wird schlechter:
  - ▶ Statt das Verstehen eines Stoffes zu fördern, werden Testaufgaben eingeübt
  - ▶ Das Curriculums wird auf getestete Sachgebiete eingengt
- Lernen wird erschwert:
  - ▶ Die Selbstachtung von Schülern und Lehrern sinkt
  - ▶ Die Verantwortung für das eigene Lernen wird vermindert
- Sanktionen führen nicht zu mehr Lernen, sondern zu mehr Betrug
  - ▶ Die Tests werden unscharf und müssen in immer kürzeren Abständen ersetzt werden.



# Sanktionen verhindern Lernen

## Schülerleistungen in US-Bundesstaaten high- vs. low-stakes testing

| Prozentsatz der US-Bundesstaaten mit...                     | Staaten mit <b>hohen Sanktionen</b> | Staaten mit <b>niedrigen Sanktionen</b> |
|-------------------------------------------------------------|-------------------------------------|-----------------------------------------|
| <b>Mathematik-Leistung über</b> dem nationalen Durchschnitt | 17 %                                | 64 %                                    |
| <b>Mathe-Leistung unter</b> dem nationalen Durchschnitt     | 67 %                                | 21 %                                    |
|                                                             | 100 % *<br>n=18                     | 100 % *<br>n=18                         |

Quelle: P. Sacks, 1999

# Die wachsende Unschärfe von Tests führt zu Schein-Gewinnen

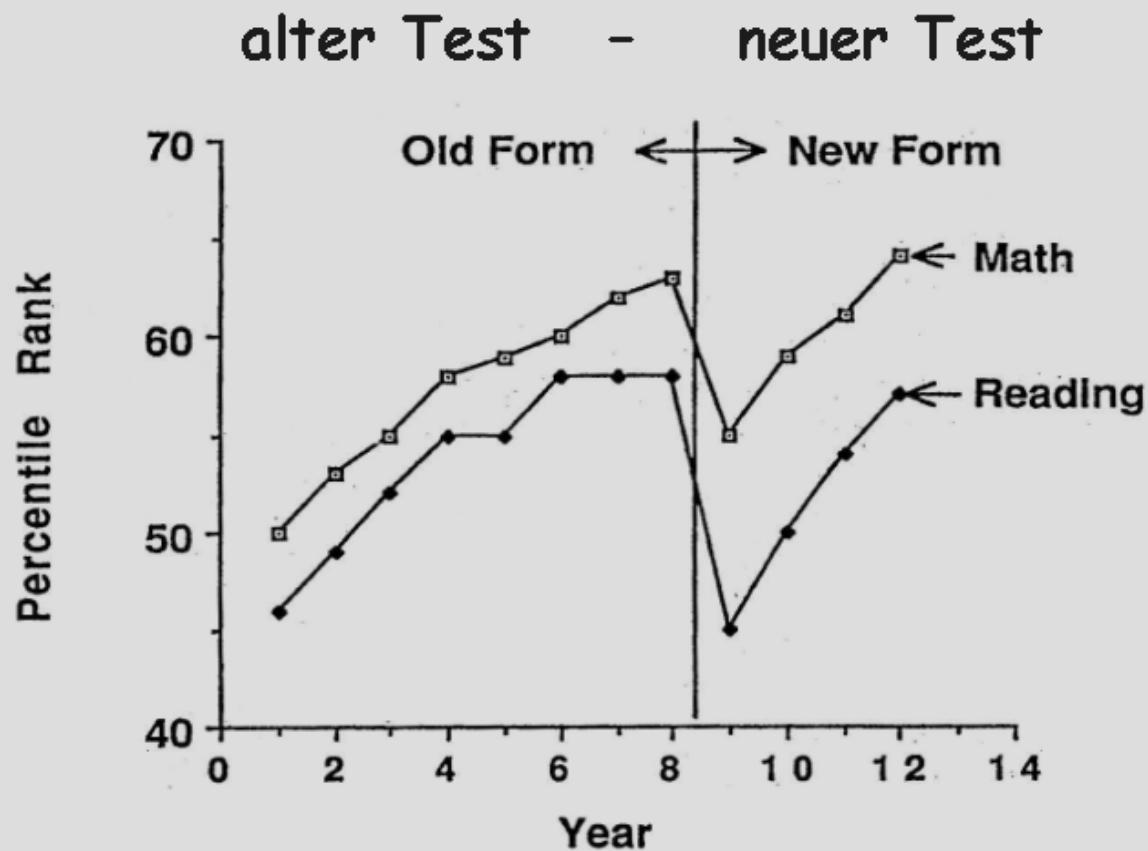


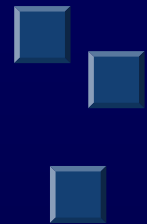
FIGURE 2. Trends in percentile rank of state means. Based on Linn, Graue, and Sanders (1990).

Linn, R., Graue, E., & Sanders, N. (1999). Comparing state and district test results to national norms: The validity of claims that 'everyone is above average.' *Educational Measurement: Issues and Practice*, 9, 5-13.



# Wirkung der Bildungsstandards in den USA

- Die Zahl der Jugendlichen ohne Schulabschluss steigt
  - ▶ 2006: 30 % (Times Magazine)
- Die Zahl der Sitzenbleiber-Rate nimmt zu.
- Qualifizierte Lehrer verlassen “in Scharen” Schulen mit niedrigen Testwerten.
- Die Qualität des Unterrichts wird schlechter statt besser:
  - ▶ Das Curriculum verarmt (“Teaching to the test”), wichtige Fähigkeiten werden vernachlässigt oder unterminiert.
- Die Leistungen in Schulleistungstests fallen (Amrein & Berliner, 2002).
- Die Kluft zwischen den Bildungschancen von Kindern aus wohlhabenden und aus armen Familien nimmt zu.





# Ruf nach einem Moratorium

“Es gibt keine überzeugenden Belege, dass der Druck, der mit Sanktions-bewehrten Tests verbunden ist, mit irgendwelchen positiven Auswirkungen für die Lernleistung der Schüler verbunden ist.

Die Belege rufen nach einem Moratorium für politische Entscheidungen, die das Bildungssystem zwingen, sich ganz von sanktions-bewehrten Tests abhängig zu machen.”

Nichols et al. (2005). High-Stakes Testing and Student Achievement: Problems for the No Child Left Behind Act. EDUCATION POLICY STUDIES LABORATORY, Sept. 2005 (meine Übersetzung, GL).

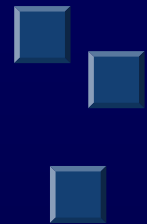
<http://www.asu.edu/educ/eps/EPRL/documents/EPRL-0509-105-EPRL.pdf>





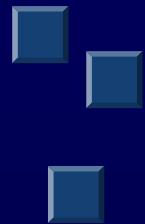
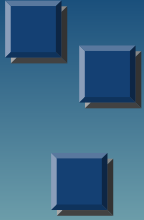
## ... nur die Testindustrie blüht

- Inzwischen hat sich ein global operierte, hochprofitable Testindustrie entwickelt (Flitner, 2006).
- In den USA mehren sich die Anzeichen von Korruption zu Lasten der Bürger und besonders der Kinder
  - ▶ Gerald Bracey (2005) zeichnet den Weg des Geldes nach, der in den USA in der Testindustrie verdient und in Lobby-Arbeit umgesetzt wird: Je mehr Tests verlangt werden, umso besser geht es ihr.
  - ▶ Der Schulrat von Houston, Texas, Rod Paige, wurde von Präsident Bush zum Bildungsminister der USA gemacht, weil es ihm besonders gut gelang, die Testwerte nach oben zu bringen – indem er tausende von test-schwachen Schülern aus der Schule gedrückt hat



# “Schule ist kein Wirtschaftsbetrieb”

Frankfurter Erklärung von 300  
Pädagogikprofessoren und Pädagogen in  
Deutschland, 10. 10. 2005



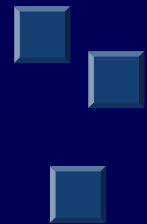
# Welche Alternative haben wir?

Ansätze zu einer demokratischen Qualitätsentwicklung  
im Bildungsbereich



# Gute Evaluation ...

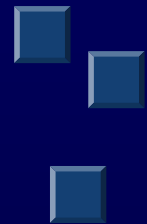
- beruht auf einer demokratischen Willensbildung (Demokratie- und Rechtsstaatsprinzip): Lehrer, Eltern und Schüler werden beteiligt.
- ist in eine Qualitätsentwicklung von Maßnahmen und Methoden eingebettet.
- wird von Belohnungen und Bestrafungen von Personen frei gehalten.
- ist aus fundierten psychologischen und didaktischen Theorien abgeleitet.
- wird ständig auf ihre Kompatibilität mit demokratisch bestimmten Bildungszielen hin überprüft.





# Demokratisches Menschenbild

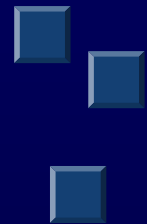
- Von welcher gesellschaftlichen Aufgabe der Schule gehen wir aus?
  - ▶ Gute Untertanen,
  - ▶ Öffentliche Reputation in Vergleichstests oder
  - ▶ Demokratie und allgemeiner Wohlstand?
- Welche Ziele hat Schulbildung?
  - ▶ Beherrschung des Bildungskanons,
  - ▶ Lösung standardisierter Testaufgaben oder
  - ▶ Erreichung von sozial *und* individuell bestimmten Lernzielen?





# Moderne Lerntheorie

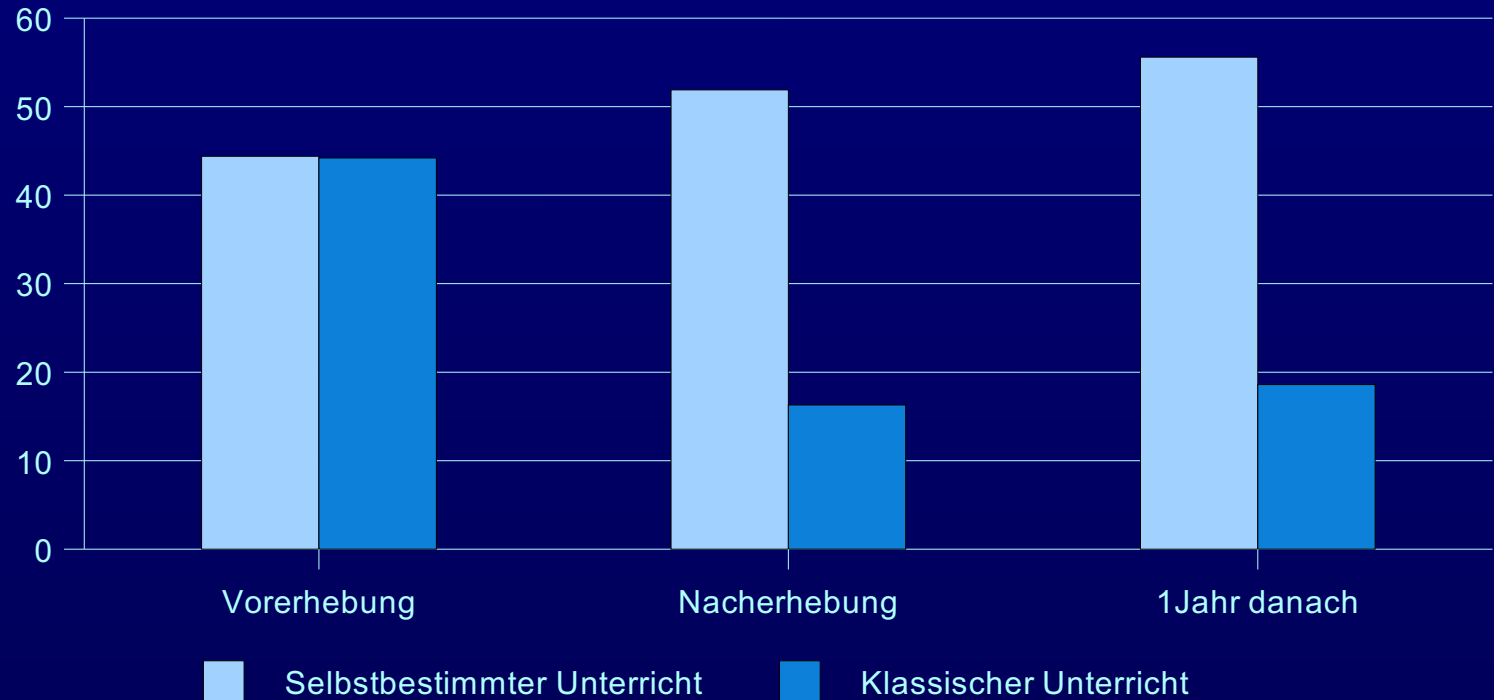
- *Passiv Lernende*: Lernen durch Druck & extrinsische Motivation (Standards, Wettkampf & Sanktionen)
- *Aktiv Lernende*: Lernen durch intrinsische Motivation, stimulierenden Lernumwelten und Interaktion mit Anderen



# Freies Lernen: Mehr Intrinsische Motivation

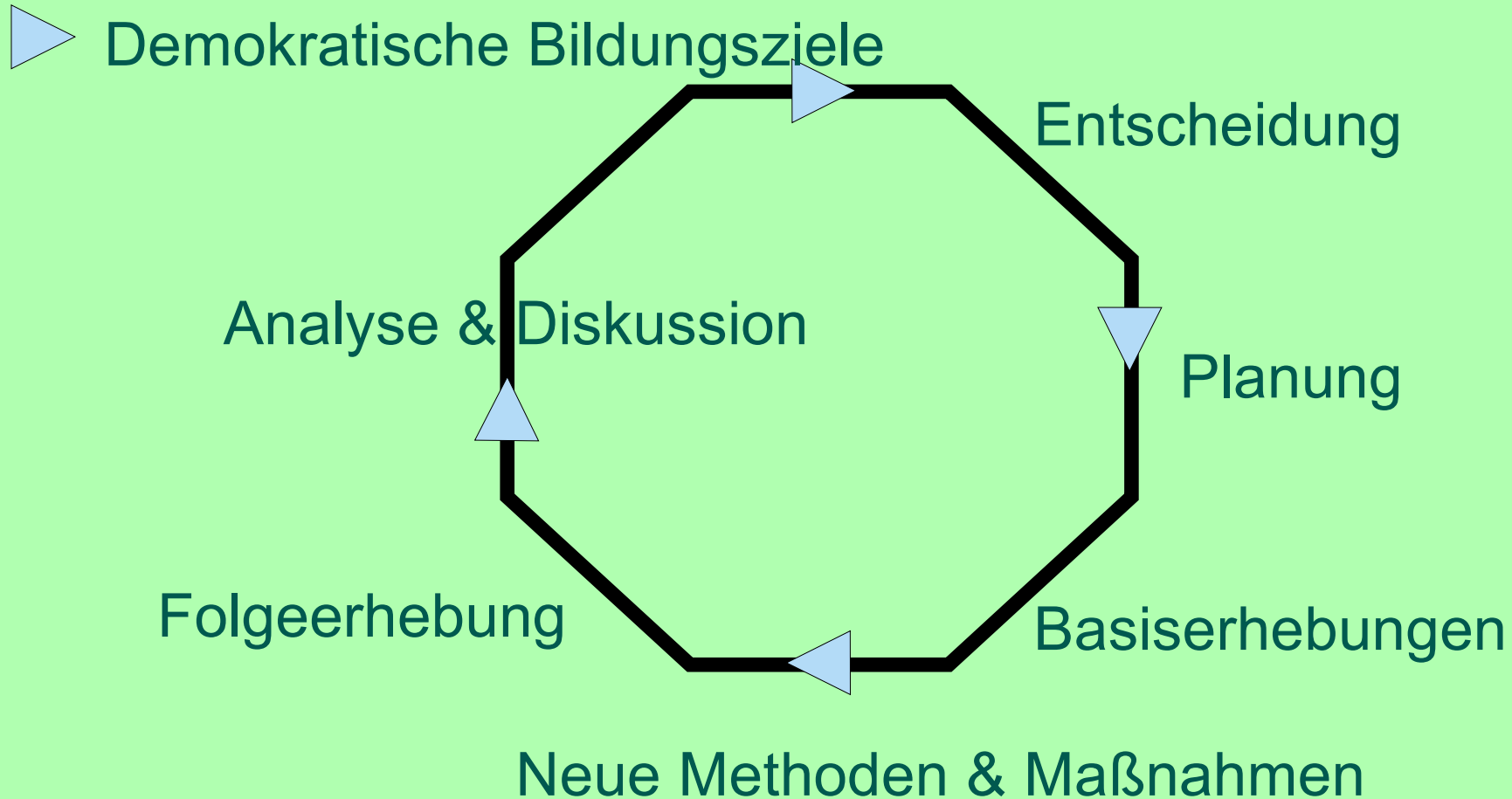
Anteil in % der  
Schüler mit  
Interesse an  
dem Stoff, der  
unterrichtet  
wurde.

(N = 70)



Quelle: Andreas Hartinger (1998). Kann eine Öffnung von Unterricht Interesse fördern? In: H. Brügelmann, Hg., Jahrbuch Grundschule. Seelze: Friedrich Verlag, S. 67-71

# Kreis-Modell der demokratischen Evaluation & Qualitätsentwicklung





# Zwei Arten der Evaluation

- Programm-Evaluation
  - ▶ Programme
  - ▶ Maßnahmen
  - ▶ Methoden
- Personen-Evaluation
  - ▶ Personen
  - ▶ Institutionen (zB. Schule)
  - ▶ Länder (zB. Schulsystem)

— Verbesserungen,  
Qualitätsentwicklung

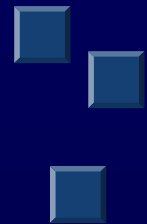
— Selektion, Sanktion,  
Gratifikation





# Ein Anwendungsfeld

- Entwicklung einer effektiven Unterrichtsmethode zur Förderung moralisch-demokratischer Kompetenzen





## Effektivere Förderung von moralisch-demokratischen Fähigkeiten im Unterricht

### ■ Vorgehen

- ▶ 1974 bis heute: Erstellung und Validierung eines Tests, der eine valide Messung moralisch-demokratischer Fähigkeiten erlaubt: Moralisches Urteil Test
- ▶ 1985 bis heute: Adaption und Weiterentwicklung einer Unterrichtsmethode (Diskussion moralischer Dilemmas; Kohlberg & Blatt).
- ▶ 1992 bis heute: Regelmäßige Evaluierung durch Vorher- Nachher-Messungen.
- ▶ Identifikation der wirksamen Bestandteile; >Flexibilisierung

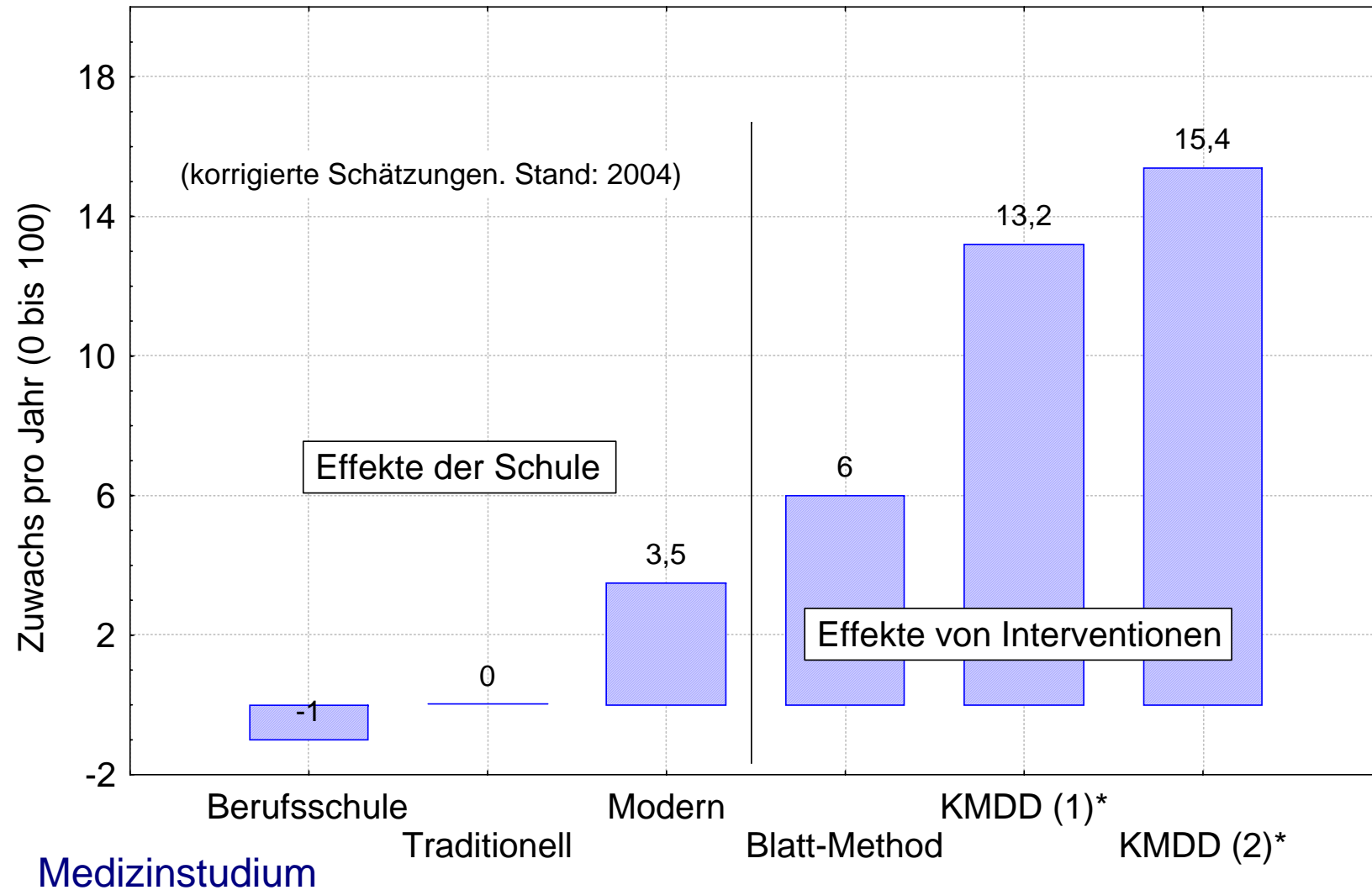
### ■ Ergebnis

- ▶ Die hocheffektive Konstanzer Methode der Dilemma-Diskussion (KMDD)
- 

Quelle: Lind (2003)

# Qualitätsentwicklung von Methoden

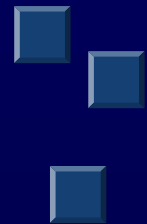
## Die Konstanzer Methode der Dilemmadiskussion (KMDD)





# Internet-gestützte Verbesserung des Unterrichts durch wissenschaftliche Selbst-Evaluation

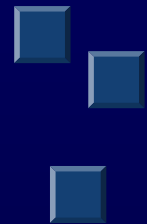
- Internet-gestützte Befragung und Testung
- Ziel-angepasste, individualisierte Messung
- Hoher Entwicklungsaufwand, geringe Anwendungskosten
- [Http://www.uni-konstanz.de/itse/](http://www.uni-konstanz.de/itse/)
- Weitere Information:
  - [Georg.Lind@uni-konstanz.de](mailto:Georg.Lind@uni-konstanz.de)





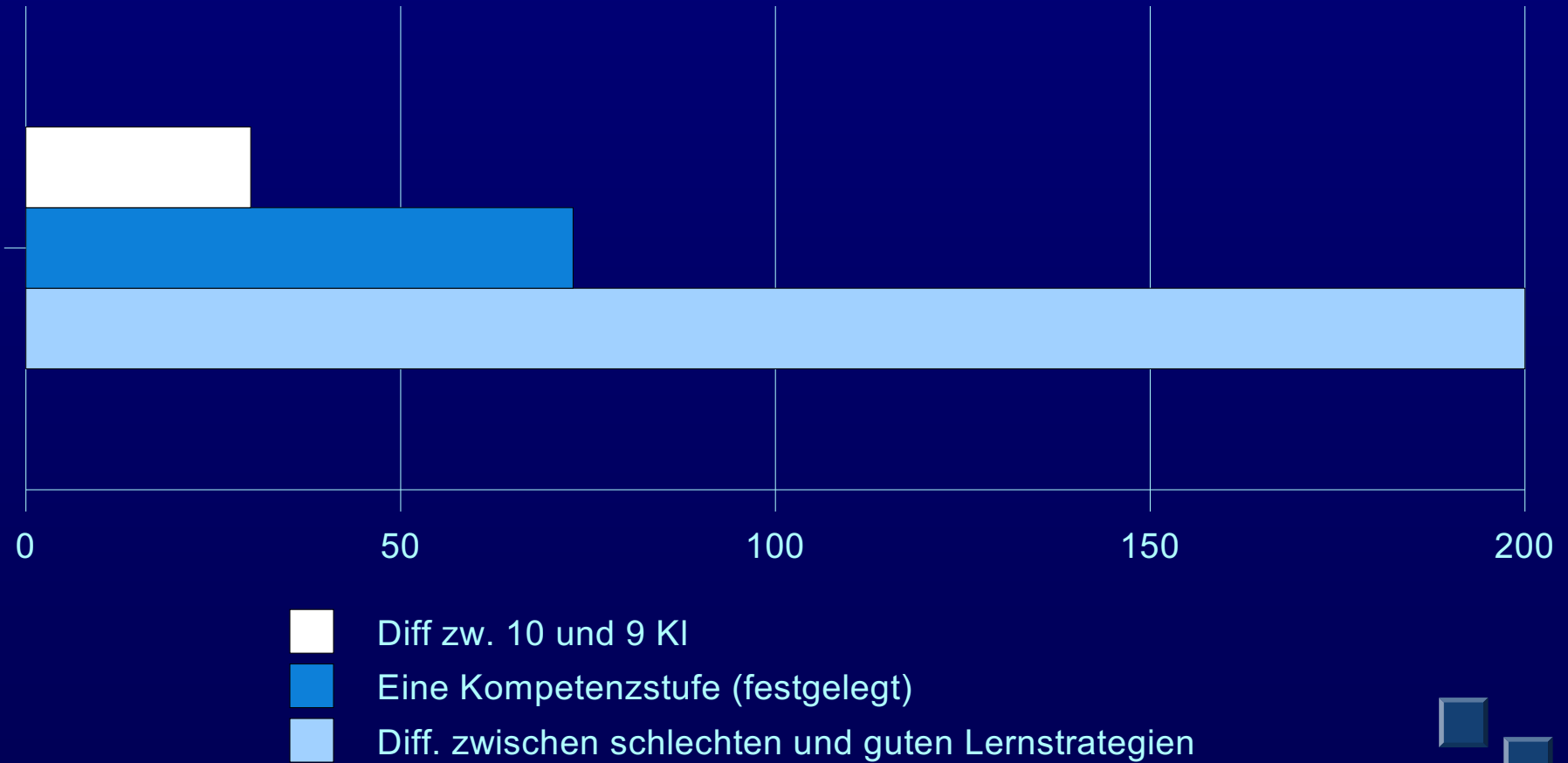
# Resümee:

- Das Lernen sollte auf allen Ebenen, top-down evaluiert werden,
  - ▶ Bildungspolitik
  - ▶ Kultusverwaltung
  - ▶ Schulverwaltung
  - ▶ Unterrichtsmethoden
  - ▶ Lernmethoden
- Aber, die Evaluation muss sich an bewährte wissenschaftliche Standards halten:
  - ▶ Maßnahmen- und Methoden-Evaluation (ohne Sanktionen)
  - ▶ Input-Output-Vergleiche
  - ▶ Transparenz und Partizipation



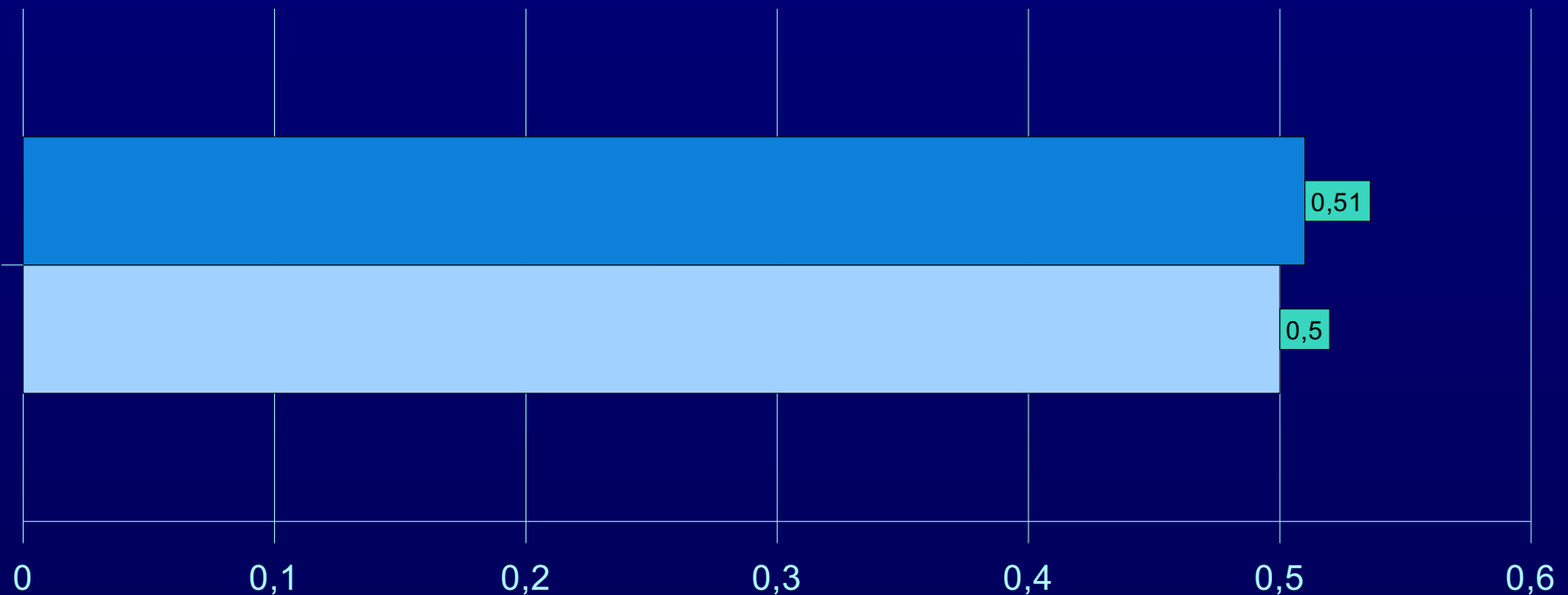
# Bessere Schule statt mehr

Unbekannte Ergebnisse von PISA 2000



# Wirkliche Schul-Autonomie hilft

Korrelation zwischen PISA-Testwerten und...



- Freiheit der Schulen, über Fächerangebot mitzuentcheiden
- Beteiligung der Lehrkräfte an der Verwaltung der Schule


Korrelation auf Länderebene



# Nachwort

Ich glaube in der Tat, dass die permanente Benotung und Bewertung von Menschen, und die damit verbundene Angstmache, die weit über das hinaus geht, was sachlich gerechtfertigt ist (zum Beispiel bei der Führerscheinprüfung oder der Approbation zum Arzt), nicht nur unsinnig und unwirksam ist, sondern einen groben Verstoß gegen die Menschenwürde darstellt.

Diese Dauerbenotung wird nur deshalb nicht auf den ersten Blick als ein Verstoß erkannt, weil sie als “Evaluation” oder Qualitätsentwicklung getarnt wird, was sie nicht ist. Bei einer Evaluation werden Methoden und Maßnahmen analysiert und bewertet, aber keine Menschen oder Menschengruppen.





# Ausgewählte Literatur

Amrein, Audrey L. & Berliner, David C. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives*, 10 (18), March 28, 2002.

Berliner, D.C. & Biddle, B.J. (1995). *The manufactured crisis. Myths, fraud, and the attack on America's public schools.* Reading, MA: Addison-Wesley.

Bracey, G. (2005). *No Child Left Behind: Where Does the Money Go?* Education Policy Studies Laboratory. EP5L-0506-114-EPRU. <http://edpolicylab.org>.

Brügelmann, H. (2005). *Schule verstehen und gestalten.* Lengwil, CH: Libelle.

Flitner, E. (2006). *Pädagogische Wertschöpfung. Zur Rationalisierung von Schulsystemen durch public-private partnerships am Beispiel von PISA.* In: J. Oelkers et al., Hg., *Pädagogische Rationalität. Studien im Umkreis Max Webers.* Zürich (im Erscheinen).

Herrmann, U. (2005). *Die nationale Testservice-Agentur IQB: der Abgesang auf pädagogische Schulentwicklung.* Bezug: <http://forum-kritische-paedagogik.de/>

Kohn, A. (1999). *Punished by rewards. The trouble with gold stars, incentive plans, A's, praise, and other bribes.* Boston: Houghton Mifflin.

Kohn, A. (2001). *Beware of the standards, not just the tests.* *Education Week*, Sept. 26, 2001

Lind, G. (2006). *Perspektive "Moralisches und demokratisches Lernen".* In: A. Fritz, R. Klupsch-Sahlmann & G. Ricken, Hg., *Veränderte Kindheit - zu pessimistisch in die Zukunft? Lehren und Lernen in der Grundschule (Arbeitstitel).* Weinheim: Beltz (in press)

Lind, G. (2004). *Jenseits von PISA - Für eine neue Evaluationskultur, S. 1 - 7.* In: Institut für Schulentwicklung PH Schwäbisch Gmünd, Ed., *Standards, Evaluation und neue Methoden. Reaktionen auf die PISA-Studie.* Baltmannsweiler: Schneider Verlag Hohengehren.

Lind, G. (2004). *Erfahrungen mit Standards in den USA – eine Übersicht.* In: U. Hameyer & H.-G. Rolff, Hg., *Standards. Journal für Schulentwicklung*, Heft 4/2004, 55-60.

Linn, R.L. (2000). *Assessment and Accountability.* *Educational Researcher*, Vol. 29, No. 2, pp. 4–16

Nichols, S.L. & Berliner, D. C. (2005). *The Inevitable Corruption of Indicators and Educators Through High-Stakes Testing.* EP5L-0503-101-EPRU, <http://edpolicylab.org>

Sacks, P. (2000). *Standardized minds. The high price of America's testing culture and what we can do to change it.* Cambridge, MA: PerseusPublishing

