From "input" to "output":

Some problems with introducing achievement standards and competency testing into the German education system

INEDD-Lecture

Siegen University, 7th Dec, 2004

by

Hans Brügelmann

The situation

The educational scene in Germany is dominated

- by a short-winded discussion of the results of international comparative studies of student achievement such as TIMSS and PISA;

- by the simplistic idea that schools would become better if we moved from an „input" to an „output" system of school management.

My main message

- There is no need to replace our input system by an output system.
- In particular, there is no empirical evidence on which to base decisions about „minimal standards" to be achieved by *all* students at the *same* time.
- The costs and side-effects of introducing high-stakes testing will outweigh the advantages.

What are the limits of international studies like PISA?

- PISA et al. can not give a comprehensive account of how good or bad our schools actually are.
- PISA et al. can not explain the reasons for the strengths and weaknesses of various schooling systems.
- PISA et al. can not specify measures to improve the quality of instruction in individual schooling systems.

What are the pay-offs of PISA et al.?

Considered in a positive light, comparative studies can be useful for the following two reasons:

- The comparative findings provide a substantial *heuristic* aid in locating problems – and in searching for explanations of and solutions to such problems.

- The prominence of the experts and the political status of the project guarantee public interest in educational policy issues, a prerequisite for inducing change into such a static system as the school.

Core curricula, educational standards, and competency tests are intended to...

- disencumber syllabi and focus them on fundamental learning objectives;
- standardize instruction in order to achieve more equity;
- improve instructional quality;
- guarantee a common basic education for all young people;
- ensure minimum levels of student achievement in major subjects;
- evaluate achievement in the various subjects in an objective and more differentiated way.

This approach is doomed to fail

...because of a threefold expectation overload:

- too many goals with partly conflicting demands are combined;
- the steering power of central management is overrated;
- the intrinsic rather than merely instrumental value of the quality of classroom processes is neglected.

A popular allegation: The German input system has failed

- Assumption 1: German 15 years olds have performed badly in PISA.
- Assumption 2: German schools are managed by input.
- Assumption 3: Successful countries have output systems.
- Conclusio: Input management is the cause of Germany's failure in PISA.
- Forecast: Introducing an output system will lead to better results in the forthcoming OECD studies.

...and the counter argument

- German primary schools have been successful in PIRLS.

- Both primary and secondary schools are managed by input in Germany.

- Therefore the management system cannot be the cause for the PISA failure of the secondary schools.

- Moreover:

  Have German students really „failed" in the international comparisons? ➔

Leseleistungen der US-SchülerInnen
im internationalen Vergleich

Kohorte Schulanfang	Rang-platz	von XX Nationen	Studie
1934-1982	7	8	IALS
1988	2	16	IEA
1991	15	31	PISA
1998	4-12	35	PIRLS

Die relativen Plätze schwanken stark innerhalb von wenigen Jahren - obwohl die Leseleistung innerhalb der USA seit 30 Jahren sehr stabil ist.

Leseleistungen der deutschen SchülerInnen im internationalen Vergleich

Kohorte Schulanfang	Rang-platz	von XX Nationen	Studie
1934-1982	2	8	IALS
1988	12	16	IEA
1991	21	31	PISA
1998	4-12	35	PIRLS

Similar to the situation in the US there are dramatic up's ad down's within very few years that cannot be explained by changes of the system.

Thus the question arises ➔

Have German students really failed?

Tab. 1

International Comparison Ranks	IEA ´91 16 Countries Age 9	PIRLS ´02 35 Countries Age 10	PISA ´00 31 Countries Age 15	IALS ´93 8 Countries Age 16-65
D'land	12	4-12	21	2
USA	2	4-12	15	7

These cohorts entered school in ...

1934-1982	IALS	:	Germany	+
1988	IEA	:	Germany	-
1991	PISA	:	Germany	-
1998	PIRLS	:	Germany	+

… and at the same time NAEP- reading results within the US remained stable…

What counts as „failure"?

Firstly, we find tremendous differences in relative positions in the international comparative studies – depending on the respective samples, designs, and methods –

but in addition we see highly diverging estimates of how many students are really „at risk" in reading ➔

Diverging estimates of students at risk in different studies

Study	Age	Year	% „at risk“
ANALFA	16+	1980	~ 5 %
IEA	14	1991	2 %
IALS	16+	1993	14 %
PISA	15	2000	25 %
PIRLS	9	2003	10 %

Divergent criteria applied to different populations lead to widely differing estimates.

Two reasons for the divergence of estimates

Threshold values for „risk" are defined in the achievement studies without ensuring

a) their ecological validity, i.e. without matching them to the requirements in everyday situations ➔

b) their biographical validity, i.e. without empirically validating the prognostic value of prerequisites for success in forthcoming learning situations ➔

Ecological validity is missing because ...

the definition of standards neglects

- ...differences between subjective contentment and test performance
- ... the high diversity in test performance of adults who are „successful" in the same occupations.

Overlap of reading competency in different groups

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

Biographical validity is missing because...

- although students with low performance at time-1 have a higher risk of failure at time-2 than high-performing students

  -- there are more students from this group who are successful than those who fail („resiliency effect");

- although students with low performance remain low performers over time

  -- on average they gain at a similar rate as high-performers („caravan effect").

Output standards: Interim summary I

Equal achievement standards for all ...

- cannot be justified as a prerequisite for successful learning in upper grades and for survival on the job market;
- do not make sense when performance at all ages differs tremendously, e.g. by four to five grade equivalents of average development within the same classroom;
- cannot do justice to individual progress from highly differing starting points.

Additional problems at the system level

The combined implementation of output standards and high-stakes testing will lead to problems because of...

- function overload by offering standards as a panacea for multiple problems* ➔
- negative side effects as illustrated over several decades in the Anglo-Saxon countries ➔

Function overload

Output standards and testing are intended to improve

- system monitoring at policy level
- management of schools and control of teachers
- assessment of student achievement and diagnosis of individual learning difficulties

Such divergent goals cannot, however, be achieved by the same instrument.

System monitoring at policy level

System monitoring by state-wide tests could usefully complement other forms of accountability. BUT:

- At the moment centralized activities dominate evaluation leading to atrophy at other levels.
- Repeating assessment every third to fifth year would be sufficient and cost much less than yearly studies.
- Sampling rather than full scale studies could fully meet the demands of system monitoring and would put less stress on schools, teachers, and students.

Controlling schools and teachers

Standards could help to focus teaching, BUT if linked to state mandated testing may lead to...

- teaching to the test (cf. US);
- narrowing of the curriculum (cf. UK);
- increasing drop (or even pull...) out of low performing students, often from minority groups (US);
- superficial adaptation rather than real change of instruction (UK, US);
- even cheating and fraud (UK, US).

Student achievement in high-stakes vs. low-stakes systems

Prozentsatz der US-Bundesstaaten mit...	Staaten mit hohen Sanktionen	Staaten mit niedrigen Sanktionen
Mathematik-Leistung über dem nationalen Durchschnitt	17 %	64 %
Mathe-Leistung unter dem nationalen Durchschnitt	67 %	21 %

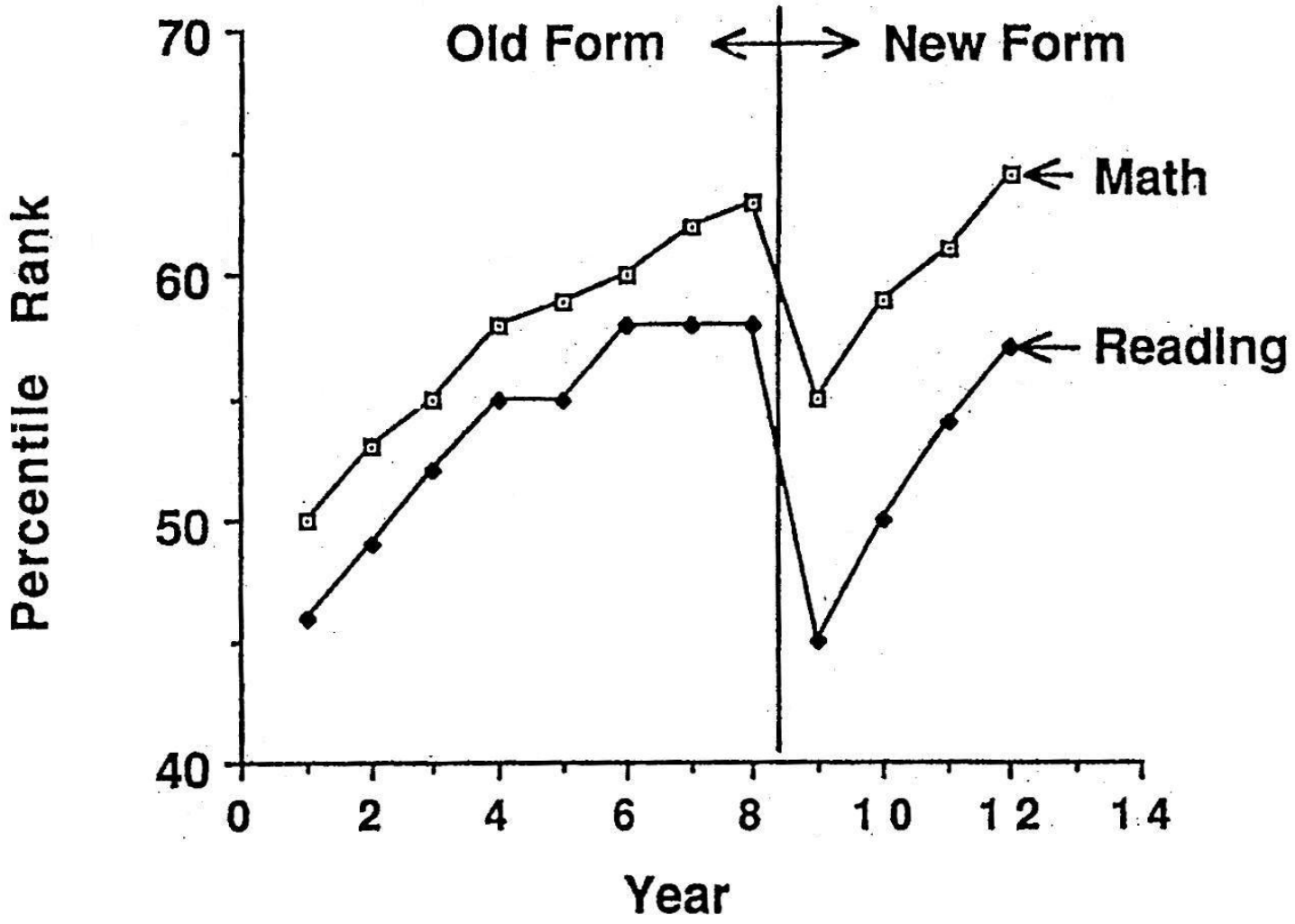

FIGURE 2. *Trends in percentile rank of state means. Based on Linn, Graue, and Sanders (1990).*

Assessment of individual learning

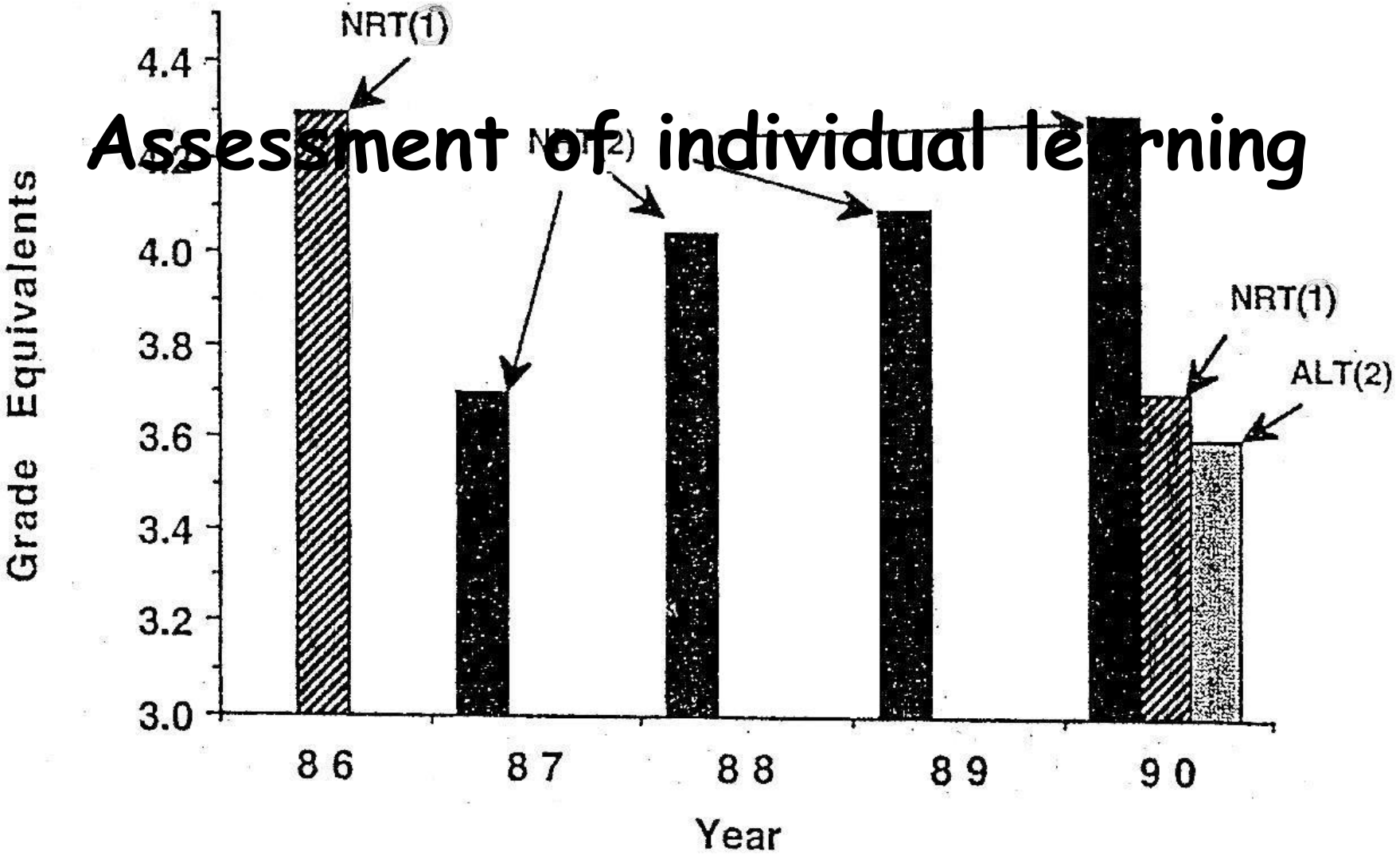

FIGURE 3. *Inflated test scores. Based on Koretz, Linn, Dunbar, and Shepard (1991).*

Competency testing: Interim Summary II

- The policy instrument „educational standards" is overstrained by too many expectations and conflicting functions
- By focussing on central control both local evaluation needs and resources are disregarded.
- Instead, the evaluation system has to be differentiated according to levels and specific functions.

From product to process

- Education will badly suffer if standards are restricted to output only.
- Such a model fosters superficial adaptation to external requirements, e.g. by teaching to the test.
- It does not improve the learning culture in the classroom – it will rather be detrimental to its development.
- Long term effects of education depend on standards for the quality of learning activities and social interaction between teachers and students.

Final summary

The „output" model of education has to be critized as being too simple and mechanistic at all levels:

- learning cannot be planned as accumulation of knowledge and skills step by step;
- classroom teaching can not „make" students learn;
- educational policy cannot manage schools by defining and controlling short term outcomes from above.

Thank you for listening...

A more detailed account can be found in:

Brügelmann, H. (2004): International tests and comparisons in education performance: A pedagogical perspective on standards, core curricula, and the quality of schooling in the German education system.

➔ www.uni-siegen.de/~agprim/printbrue.htm

to be published in:

Rotte, R. (ed.) (2005): International perspectives on education policy. Nova Science Publ.: New York (forthcoming).