The Role of Large Scale Assessments in Research on Educational Effectiveness and School Development

Eckhard Klieme

German Institute for International Educational Research (DIPF)
Frankfurt am Main, Germany

DIPF
Educational Research
and Educational Information

International Conference
“The Role of International Large Scale Assessments in Educational Policy”

ETS, Princeton, March 18, 2011

Overview

1. **LSA: Goals and limitations of (international) Large Scale Assessments**
2. **EER: Goals and design of educational effectiveness studies**
3. **EER → LSA:**

   Designing LSA based on educational theory and research

   > OTL > Quality of instruction
4. **LSA → EER:**

   How LSA may contribute to our knowledge of educational effectiveness and school development

   > Exploration > Hypothesis testing > Understanding context
5. **LSA + EER:**

   Enriched (panel) designs integrating LSA and EER:

   > Explaining student growth (German Study on Language Instruction)

   > Explaining school development (PISA School Panel Germany)
6. **Conclusions**

1. (International) Large Scale Assessments

1959

- 1959 - Pilot Twelve-Country Study
- 1964 - First International Mathematics Study
- 1970 - Study of Reading Comprehension
- 1970 - Study of Literature Education
- 1970 - First International Science Study
- 1970 - Study of French as Foreign Language
- 1970 - Study of English as Foreign Language
- 1970 - Study of Civic Knowledge
- 1980 - SIMS: Second International Mathematics Study (IEA)
- 1981 - CES: Classroom Environment Study
- 1983 - SISS: Second International Science Study (IEA)
- 1986 - PPP: Pre-Primary Project (IEA, 1986-2003)
- 1988 - International Assessment of Educational Progress (ETS): IAEP I
- 1990 - International Assessment of Educational Progress (ETS): IAEP II
- 1990 - International Reading Literacy Study
- 1994 - IALS: International Adult Literacy Survey (OECD, StatCan, ETS, UNESCO) – Wave 1
- 1995 - Third International Mathematics and Science Study
- 1996 - IALS: International Adult Literacy Survey (OECD, StatCan, ETS, UNESCO) – Wave 2
- 1996 - CivEd: Civic Education Study (IEA)
- 1997 – SITES-M1: Second Information Technology in Education Study Module 1 (IEA)

1998

1998

- 1998 - IALS: International Adult Literacy Survey (OECD, StatCan, ETS, UNESCO) – Wave 3
- 1998-2000: SACMEQ: Southern Africa Consortium for Monitoring Educational Quality
- 1999 - Trends in International Mathematics and Science Study
- 1999 - CivEd: Civic Education Study (IEA)
- 1999 – SITES-M2: Second Information Technology in Education Study Module 2 (IEA)
- 2000 - PISA: Programme for International Student Assessment (OECD)
- 2003 - PISA 2003: Programme for International Student Assessment (OECD)
- 2003 - ALL: Adult Literacy and Life Skills Survey (StatCan, ETS, OECD, UNESCO)
- 2006 - ALL: Adult Literacy and Life Skills Survey (StatCan, ETS, OECD, UNESCO)
- 2006 – SITES 2006: Second Information Technology in Education Study 2006 (IEA)
- 2008 – TEDS-M: Teacher Education and Development Study in Mathematics (IEA)
- 2009 - ICCS 2009: International Civic and Citizenship Education Study (IEA)
- 2010 - PIAAC: Programme for the International Assessment of Adult Competencies (OECD)
- 2011 – PIRLS 2011: Progress in International Reading Literacy Study 2011 (IEA)
- 2011 – TIMSS 2011: Trends in International Mathematics and Science Study 2011 (IEA)
- 2013 – ICILS: International Study of Computer and Information Literacy

2011

(graphical display with permission from Knut Schwippert, Univ. of Hamburg)

LSA provide Indicators for policy making

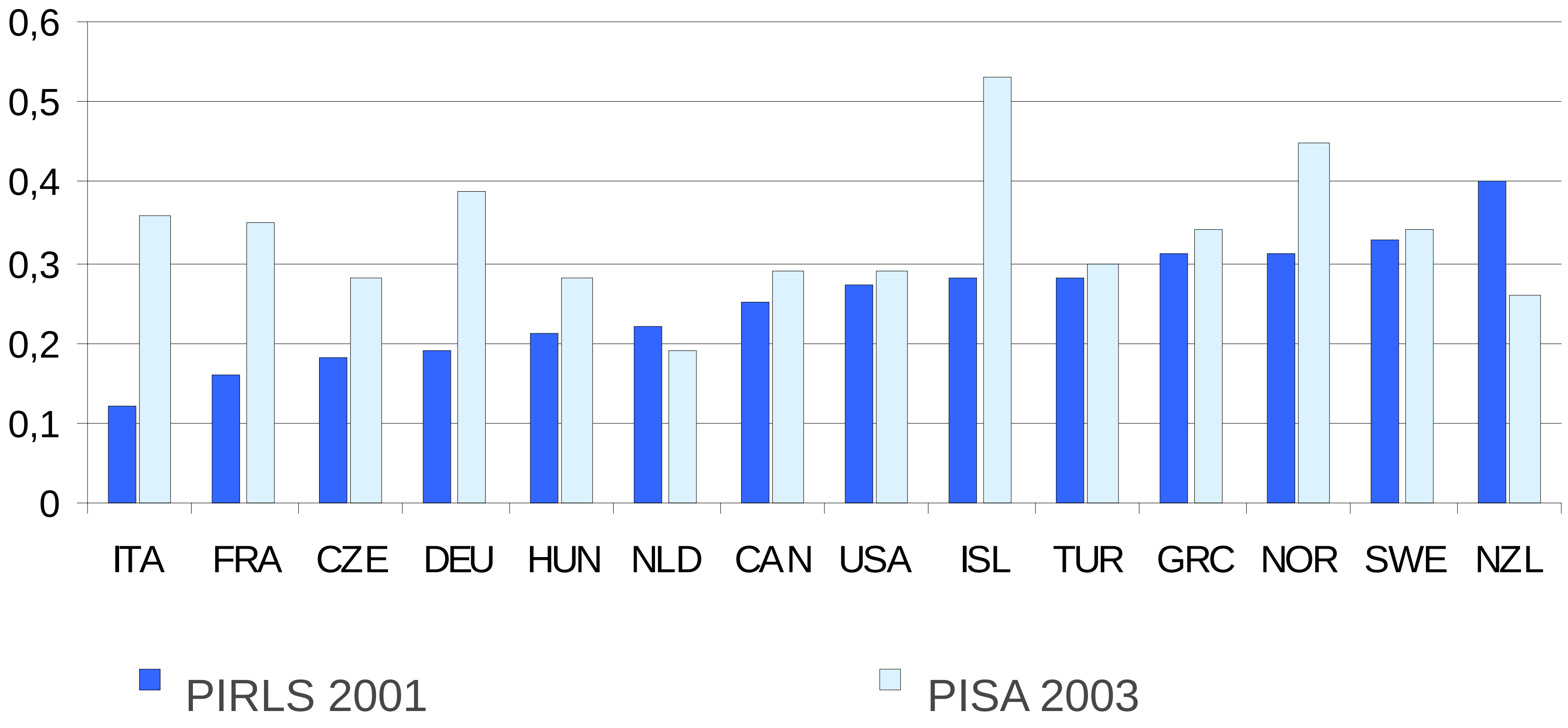
1. describing state and change of systems
2. setting and monitoring benchmarks, both inter- and intranationally
3. identifying strengths and weaknesses, success and challenge
4. communicating to the public and the professionals

With regard to

- **Functioning** of the educational system (inputs, processes)
- **Effectiveness**: Output (achievement, certificates) and/or outcome (yield: competencies, educational pathways)
- **Equity**: Distribution of input, process, and outcome by (e.g.) gender, SES, immigration background, region
- **Efficacy**: Outcome (value-added ?) per investment
- **Production function**: Relationsship between input-process-outcome at the individual, classroom/teacher, school, system level

Example: Monitoring Gender Inequity in Reading:

Gender differences in favor of girls (effect sizes, d) in PIRLS 2001 and PISA 2003



LSA databases allow for sophisticated analyses in educational research

Example:

Between-School differences in grading standards

depending on centralized vs. decentralized exit exams

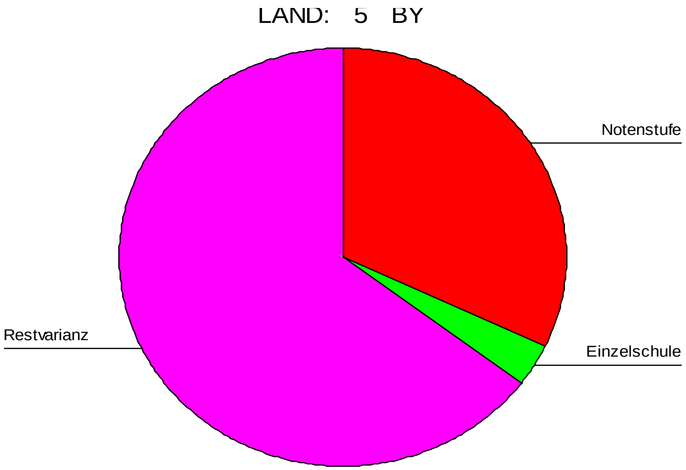
Data from German extension to PISA 2000 (about 100 schools per state)

Method: decomposing variance in PISA-outcomes (Math, reading, science)

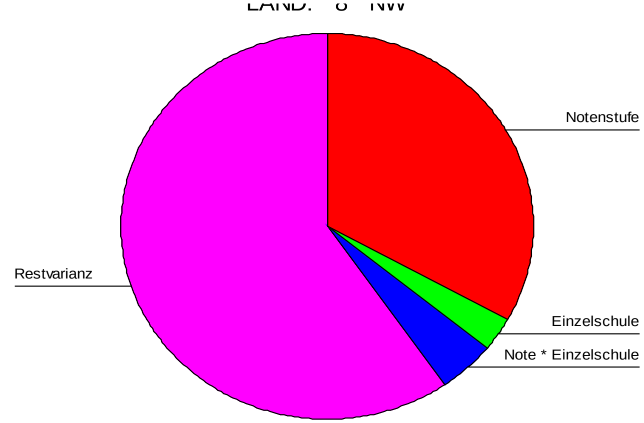
by school, grade level, school X grade level

within school types, states

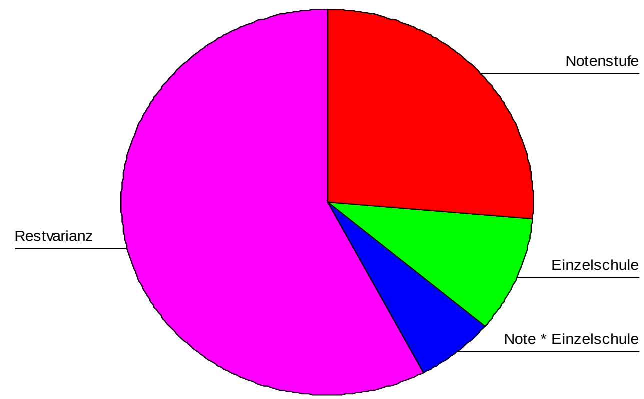
Bavaria, High track: Central exams



Northrhein-Westfalia, High track: No central exam



Bavaria, Low track: No central exam



Variance in PISA-scores attributable to

- Schools
- Grade levels (1-6)
- Interaction school X grade

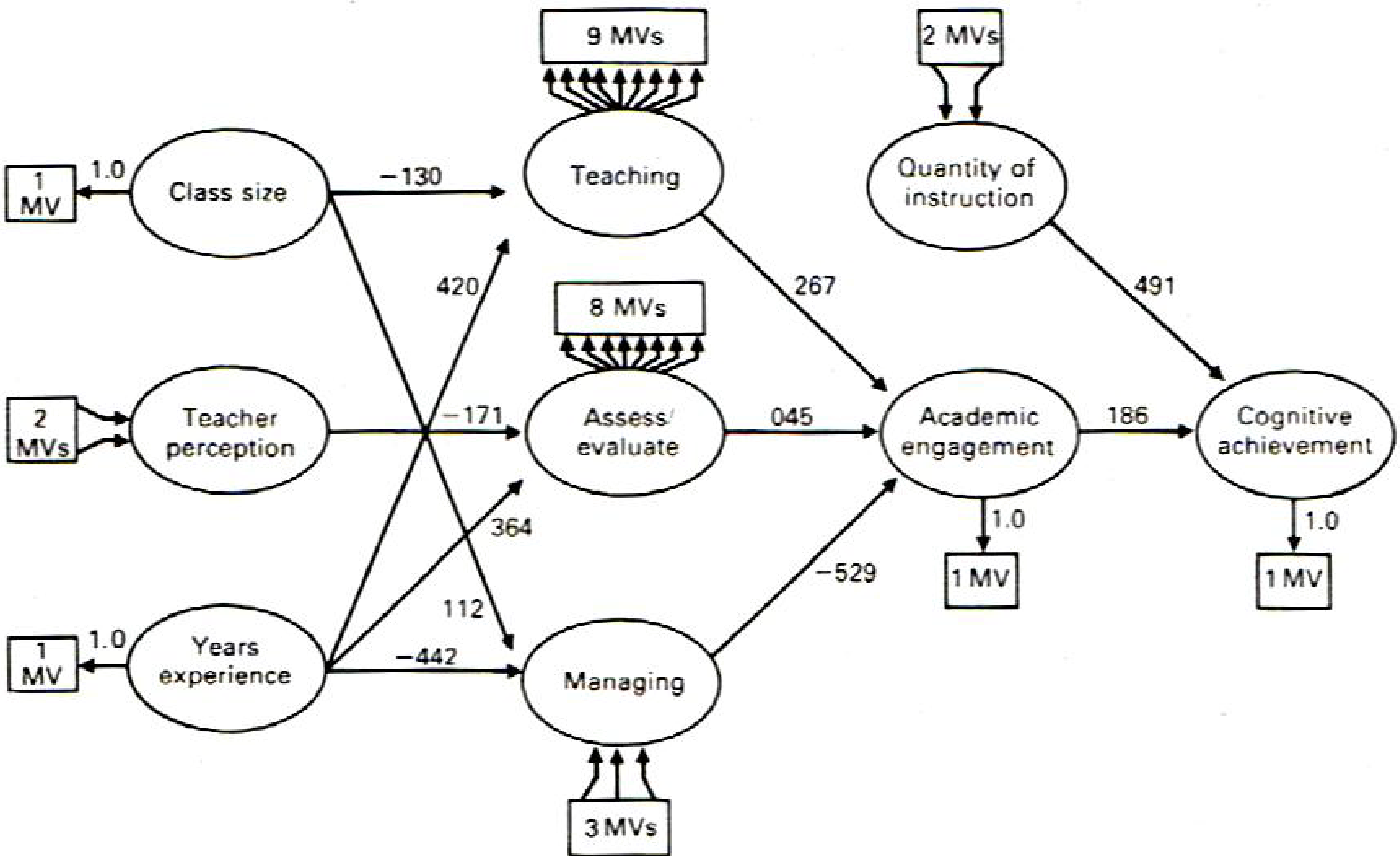
Limits of Large Scale Assessments: Reverse Causality on the classroom level

Professional educators, in their daily practice, need to monitor student development, and change treatments accordingly.

Examples: assigning tasks, forming groups for collaborative learning, giving feedback, deciding on grade retention/promotion …

→ Effects of these treatments cannot be estimated from cross-sectional data only, without knowing prior achievement and other factors that drive treatment assignment

Limits of Large Scale Assessments: Reverse Causality (IEA Classroom Environment Study, Anderson et al. 1989)

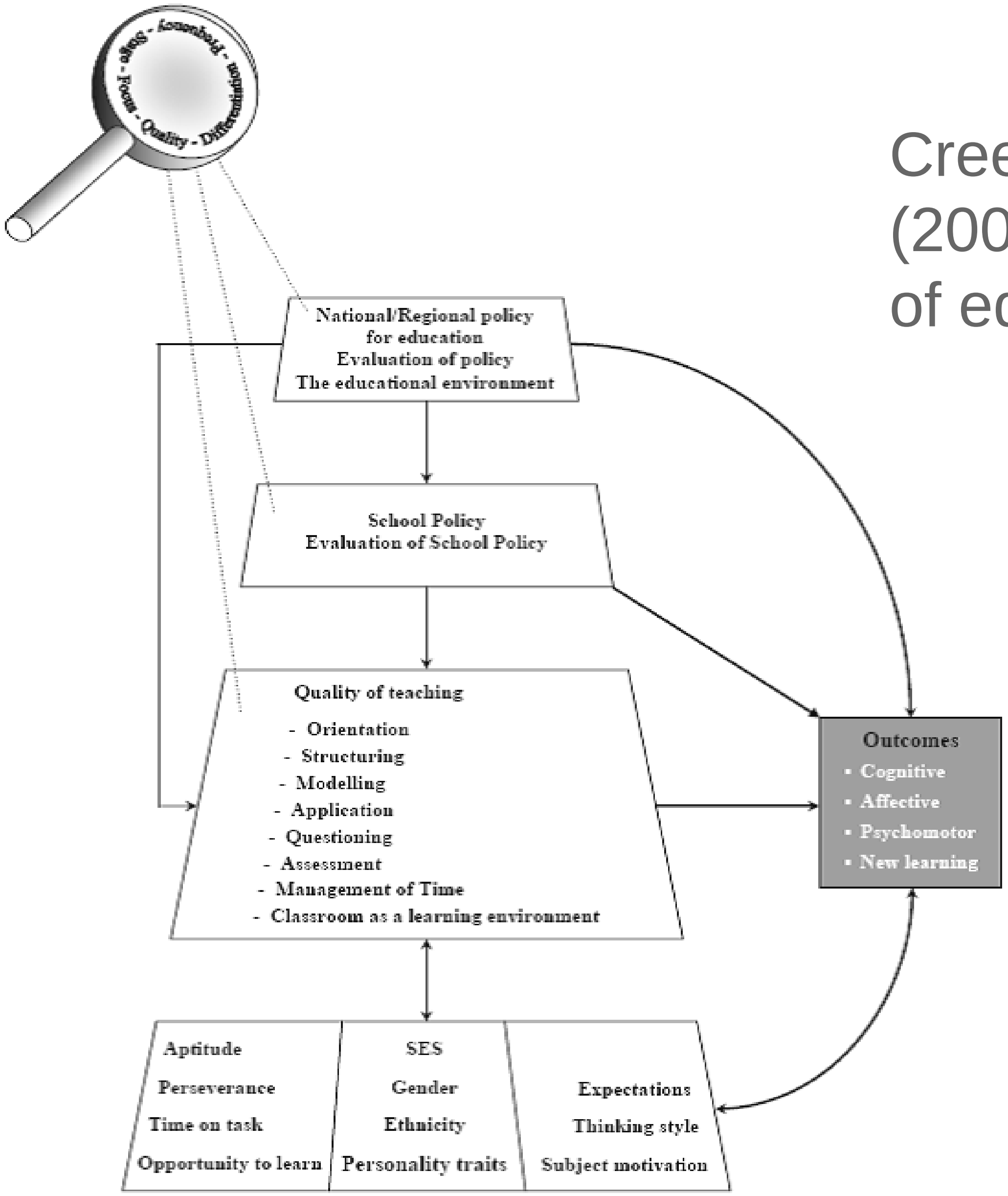


Limits of Large Scale Assessments: Reverse Causality on the School level (OECD 2007)

Box 5.5 **Multilevel models: Accountability policies**

Accountability policies and student performance	Gross		Net	
	Change in score	p-value	Change in score	p-value
School informing parents of children's performance relative to other students in the school (1=yes; 0=no)	4.7	(0.140)	2.8	(0.139)
School informing parents of children's performance relative to national benchmarks (1=yes; 0=no)	4.2	(0.100)	1.8	(0.228)
School informing parents of students' performance relative to other schools (1=yes; 0=no)	-5.0	(0.013)	-1.4	(0.352)
School posting achievement data publicly (1=yes; 0=no)	14.7	(0.000)	6.6	(0.000)
School using achievement data for evaluating principals (1=yes; 0=no)	-2.3	(0.354)	0.0	(0.993)
School using achievement data for evaluating teachers (1=yes; 0=no)	4.3	(0.076)	-0.5	(0.711)
School using achievement data for allocating resources to schools (1=yes; 0=no)	-4.8	(0.034)	-4.3	(0.007)
School with achievement data tracked over time (1=yes; 0=no)	-2.4	(0.327)	-1.2	(0.443)
System with standards-based external examinations (ratio of existence)	36.1	(0.028)	17.0	(0.226)

2. Educational Effectiveness Research (EER)



Creemers & Kyriakides (2008): The dynamic model of educational effectiveness

4 stages of Educational Effectiveness Research (Creemers, Kyriakides & Sammons, 2010)

1. Showing that „school matters“: estimates of variation
2. Searching for factors associated with student outcomes

   e.g., Edmonds (1979) 5 factors:

   leadership, expectations, emphasis on basic skills, orderly climate, evaluation

   see also Bryk et al. (2010): „essential supports“
3. Deveplomen of theoretical models
4. Analysing the complex nature of educational effectiveness;

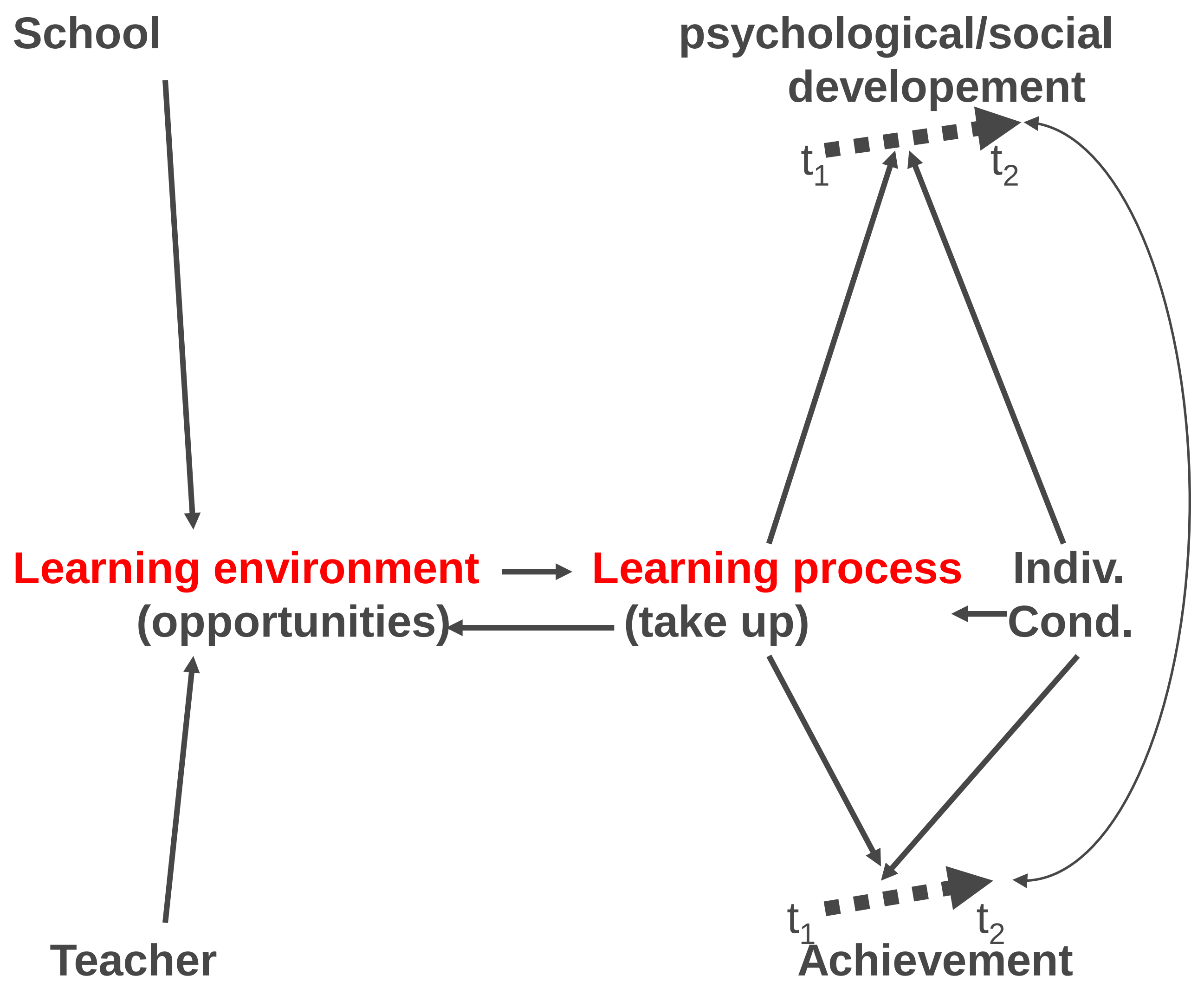
   links with the study of school improvement

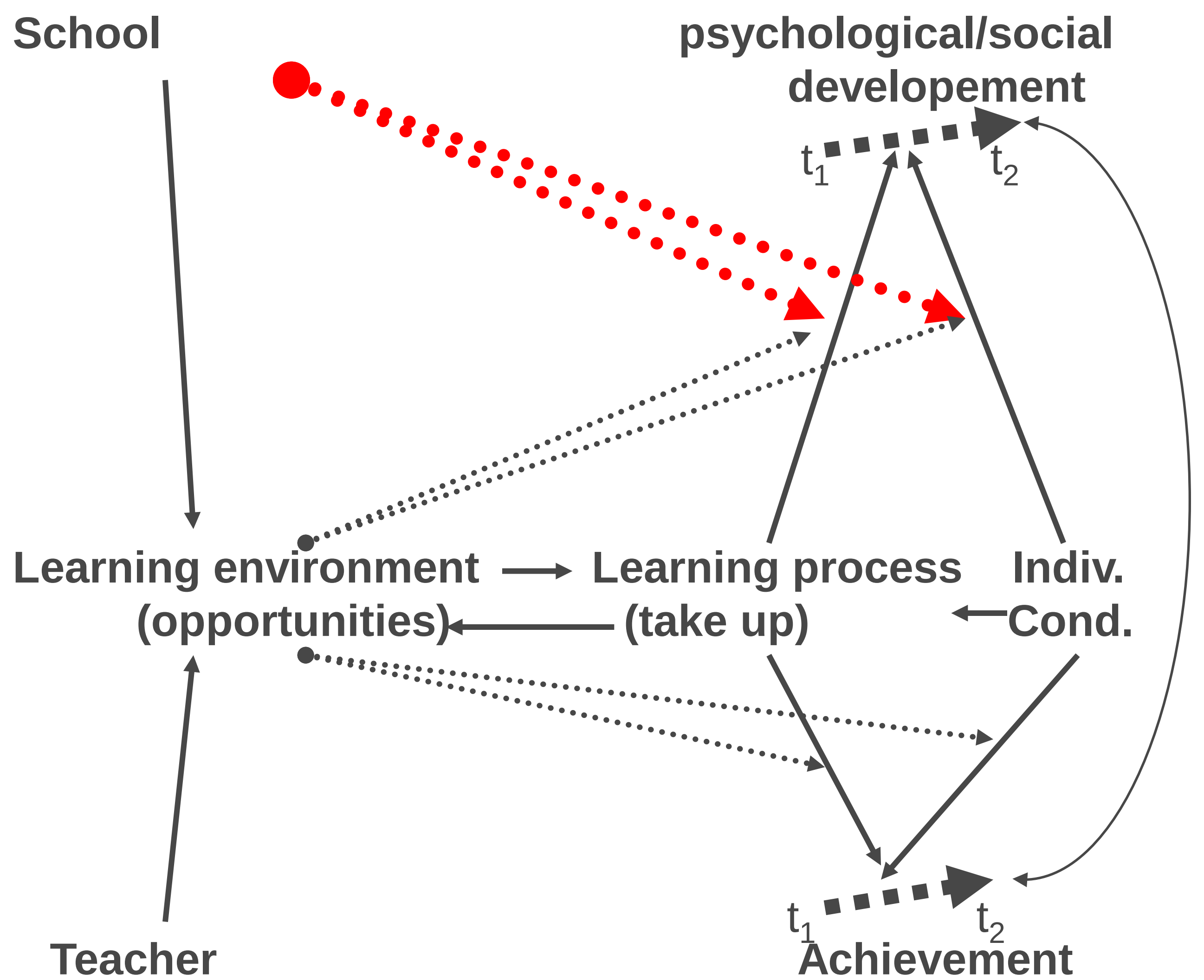
Stage 3: CIPO-Model (Scheerens/Bosker 1997)

CONTEXT
School structure, Curriculum, pedagogical traditions and orientations, teacher education, budgeting and regulation, socio-economic and cultural context

INPUT	PROCESS	OUTPUT
Teacher-student-rate, qualification of teaching staff, student population, parent commitment	Quantity of instruction, School curriculum, leadership, teacher cooperation and collaboration, professional development, cohesion, school culture (norms and values), school climate, internal and external evaluation	**School level**
Students per class, teacher competencies	Instructional quality: - opportunity to learn - clear, well-structured classroom management, - supportive, student-oriented classroom climate, - cognitive activation with challenging content	**Classroom level**
SES, social and cultural capital, family support, gender, language and migration background, general intellectual ability, pre-knowledge	time invested, self regulation, motivation and interest, self concept, learning strategies	**Individual level**

Stage 4: Mediating and moderating variables




School
psychological/social developement
t_1
t_2
Learning environment (opportunities)
Learning process (take up)
Indiv. Cond.
t_1
t_2
Teacher
Achievement

Challenges to Educational Effectiveness Research

- The adaptive nature of educational processes
- The complexity of mediating processes
- The importance of moderating variables
- The weakness of distal effects, especially school effects

Challenges to Educational Effectiveness Research

- The adaptive nature of educational processes
- The complexity of mediating processes
- The importance of moderating variables
- The weakness of distal effects, especially school effects

- The dynamic functioning of educational institutions
- The incoherence and instability of effect sizes
- The fundamental difference between status (at a given moment), individual gains, and organizational change (over time)

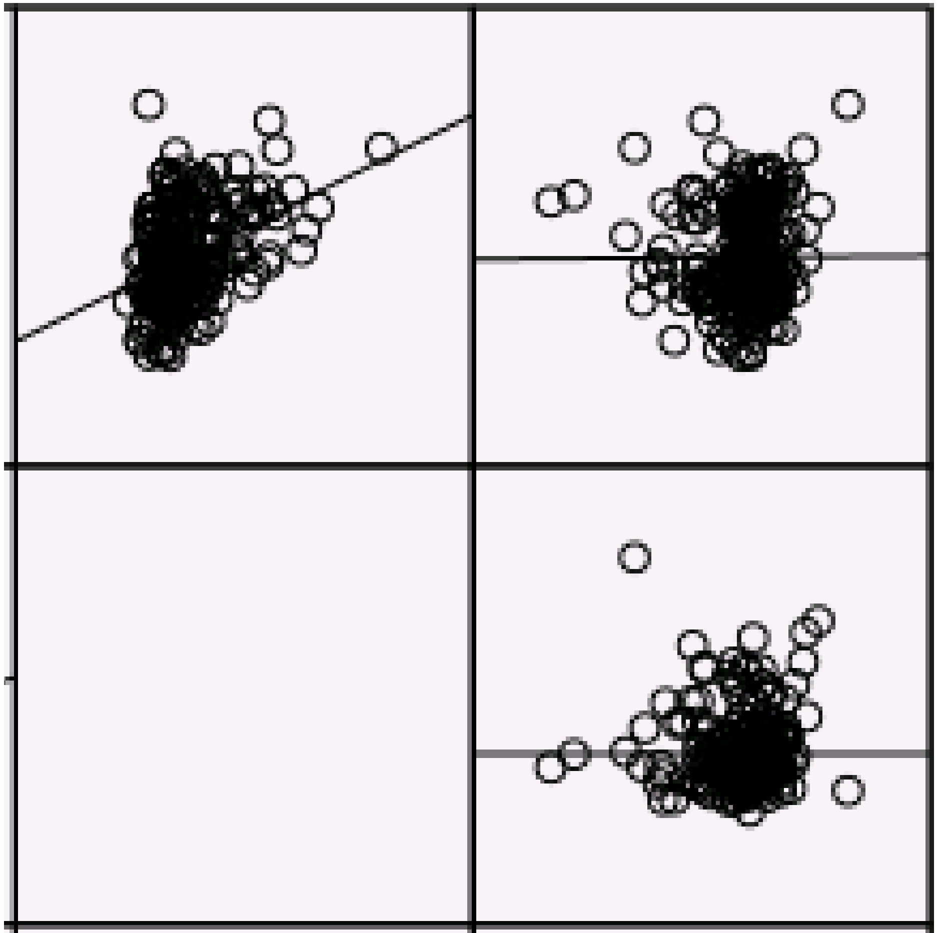
3 Types of „value-added“ measure (n=273 schools)

Mean achievement gain within 1 year

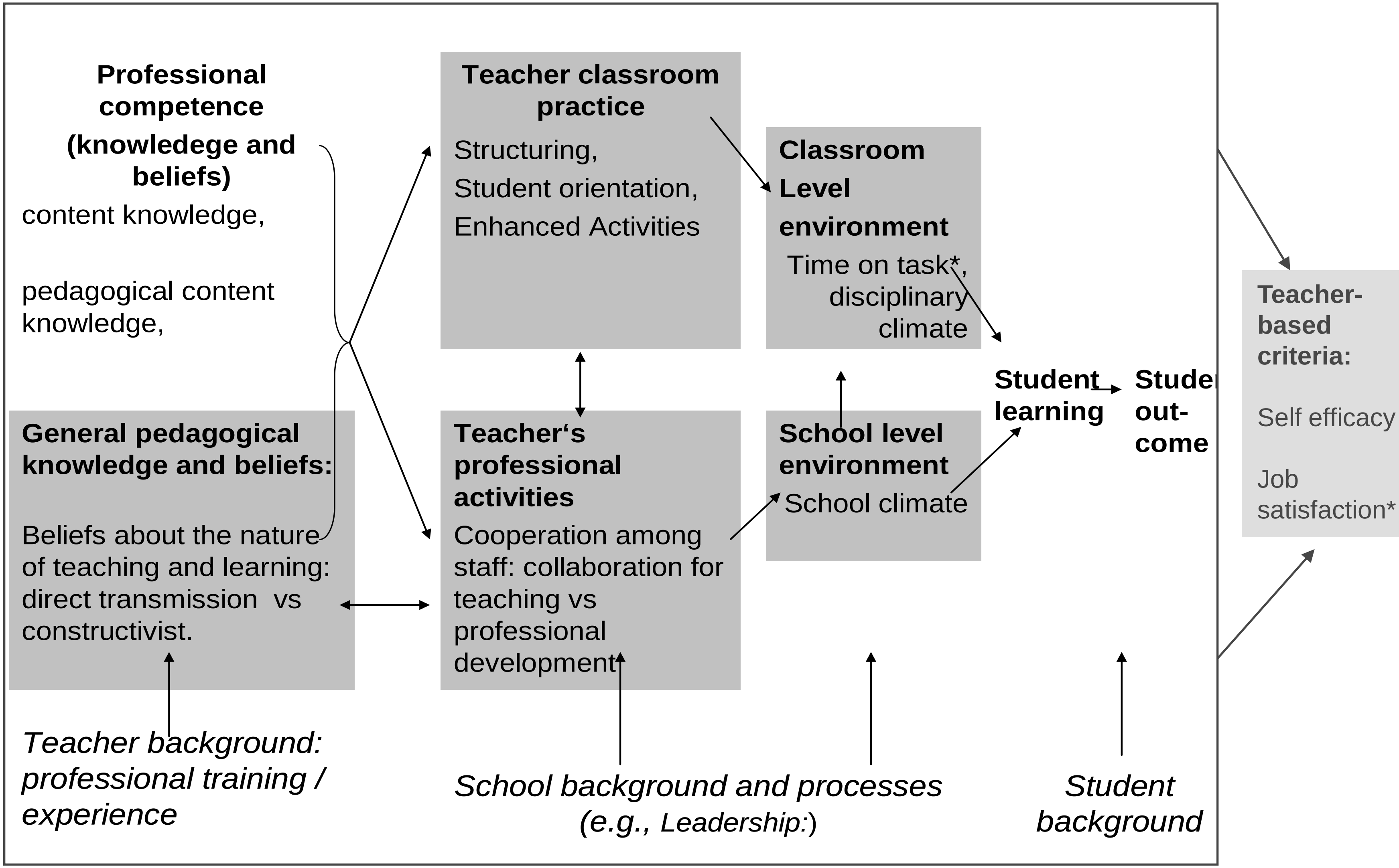
Change in mean gain over 2 yrs

Adjusted means (adjusted for SES, gender, immigration status & compositional effects):
-> Stability = .70

Mean achievement gain within 1 year:
-> Stability = .31



3. How EER can inform LSA: TALIS



OECD - TALIS: section on teacher beliefs and teaching practices

3. How EER can inform LSA: PISA

(I) General (i.e., domain-independent) trend variables

General input variables:

- Student-level inputs (grade, gender, socio-economic background= parental education and occupation/family wealth/educational resources/cultural possessions, migration data = immigration status / heritage language/ age on arrival in country, family support)
- School-level contexts and inputs (community size, resources, qualifications of teaching staff)

General process variables:

- School-level processes (decision-making, admission policies, assessment and evaluation policies, professional development, student-teacher-relations, parental involvement)
- Instructional processes (learning time, disciplinary climate, teacher support)

General outcome variables:

- General non-cognitive outcomes – Commitment to learning (behavioural: truancy; personal goal: educational aspirations; motivational: learning engagement, affective: sense of belonging)

(II) Domain-specific trend variables

- Domain-specific cognitive outcomes (math, science, reading literacy)
- Domain-specific non-cognitive outcome variables (strategies and metacognition, domain-related beliefs, self-related beliefs, motivation)
- Domain-specific process variables (Opportunity to learn, instructional quality, system- and school-level support)

(III) Thematic extension variables (extensions within individual cycles), e.g.

- International options (e.g. in PISA 2012, educational career/second language learners; ICT literacy)
- Context variables for additional domains (e.g. ICT-related experiences relevant for computer-based problem solving)

Opportunity to learn: The „observable structure“ of teaching

(Schmidt & McKnight 1995; Schmidt & Maier 2009)

- Content Coverage
- Content Exposure Variables:
  considering time and depth of teaching
- Content Emphasis Variables:
  e.g., lower vs.higher order skills
- Quality of Instructional Delivery Variables:
  classroom teaching practices

Quality of instruction: Empirical findings (Brophy,2000)

2. Opportunity to learn
4. Establishing learning orientations
5. Coherent content
11. Goal-oriented assessment

1. A supportive classroom climate
10. Co-operative learning

3. Curricular alignment
6. Thoughtful discourse
7. Practice and application activities
8. Scaffolding students' task engagement
9. Strategy teaching
12. Achievement expectations

Quality of instruction: Empirical findings (Brophy,2000)

2. Opportunity to learn

4. Establishing learning orientations

5. Co

11. G

1. A

10. G

3. Cu

6. Th

7. Pr

8. Sc

9. St

12. Achievement expectations

Opportunity to learn:

Students learn more when most of the available time is allocated to curriculumrelated activities and the classroom management system emphasizes maintaining their engagement in those activities.

(Brophy 2000, p. 10)

Quality of Instruction: Theoretical foundations

Process-Product-Research; Behavioral learning theory

→ **Classroom Management, Clarity & Structure**

Humanistic pedagogy and psychology;

Self determination Theory (Deci & Ryan)

→**Supportive climate & Structured learning environment**

Cognitive Theory (e.g. Brown 1997, Mayer 2004);

concepts from (moderate) constructivism

→ **Cognitive Activation & Deep Content**

Quality of Instruction: The „latent“ structure of teaching

Pianta & Hamre (2009): CLASS framework

- Classroom organization
- Emotional support
- Instructional support

Tschannen-Moran, M. & Woolfolk Hoy, A. (2001):

Ohio teacher efficacy scales (OSTES)

- Efficacy for classroom management
- Efficacy for student engagement
- Efficacy for instructional strategies

Quality of Instruction: The „latent“ structure of teaching

Quality of instruction (opportunities provided)	**Mediation (take up)**	**Effects**
Classroom management, clarity & structure	(Time on task)	Knowledge and understanding
Cognitive Activation & deep content	(Depth of processing, reflection)	
Supportive Climate	(self determination)	Motivation

(Klieme, Pauli & Reusser, 2009; Klieme, Schümer & Knoll, 2001)

4. How LSA can inform EER

> exploring & generating hypotheses

2nd order factors of classroom practice

based on high-inference video-ratings (Clausen,Klieme & Baumert 2002

(TIMSS-Video 1994 Germany: national sample, 100 + 86 lessons)

Classroom Management	Supportive climate	Cognitive Activation
Effective treatment of interruptions „teacher intervenes immediately, before disturbance may evolve“ Clarity of rules Interruptions (-) Waste of time (-) Monitoring Time on task Teacher Unreliability (-) **Clarity and structuredness of the Instruction**	Social orientation: „teacher takes care of his students‘ problems“ Teachers diagnostic competence with regard to social behavior **Individual reference norm in evaluation** Rate of interaction (-) Pressure on students (-)	Teacher’s ability to motivate students: „can present even abstract content in an interesting manner “ **Errors as opportunities** **Demanding tasks** Practicing by repetition (-)

4. How LSA can inform EER

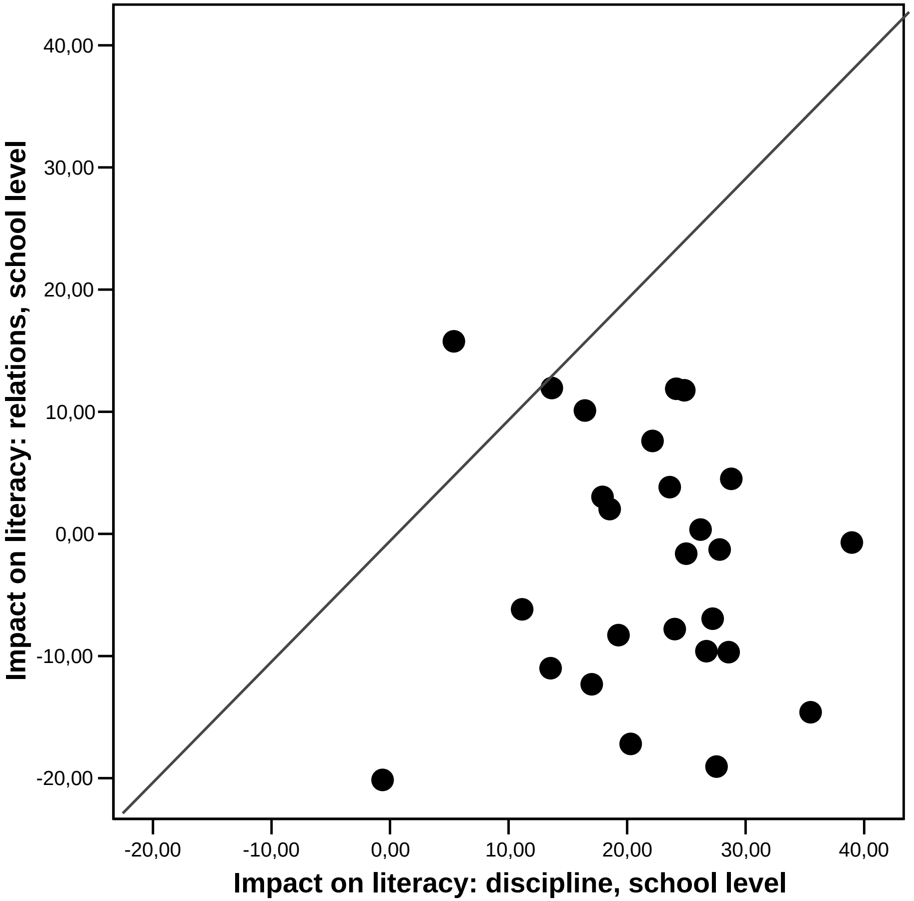
> exploring & generating hypotheses

> testing hypotheses

PISA 2000: School-level effects by country

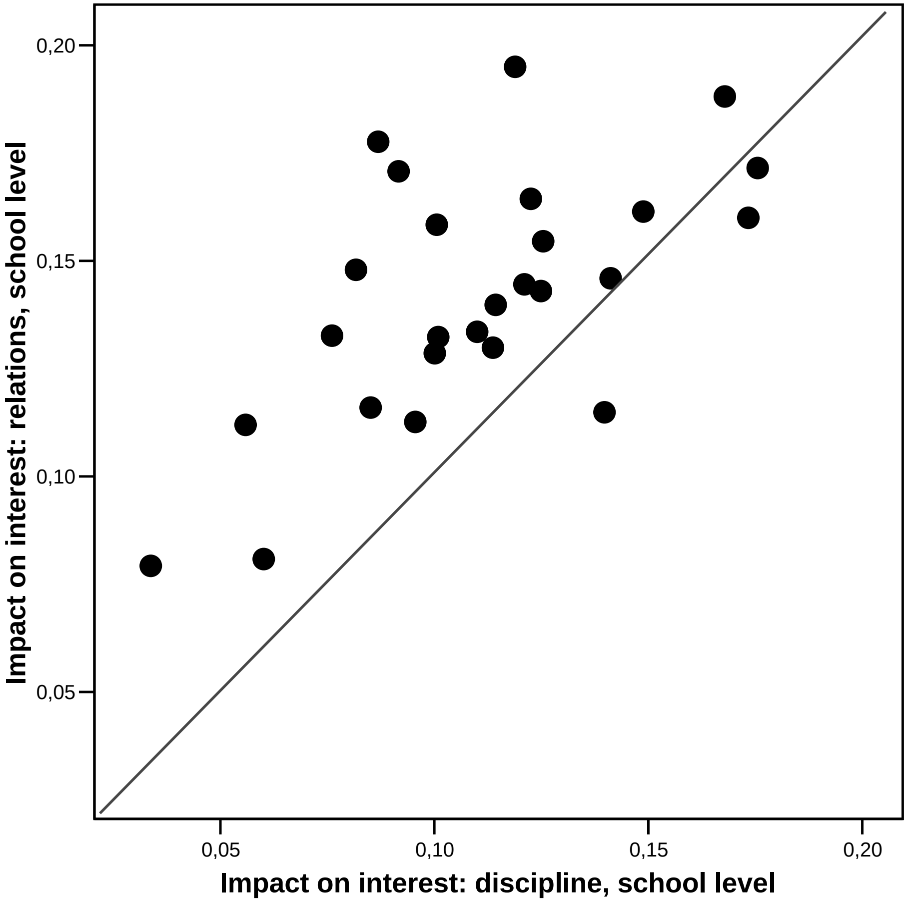
… on Reading Literacy:

Discipline > Relations



… on Interest in Reading:

Relations > Discipline



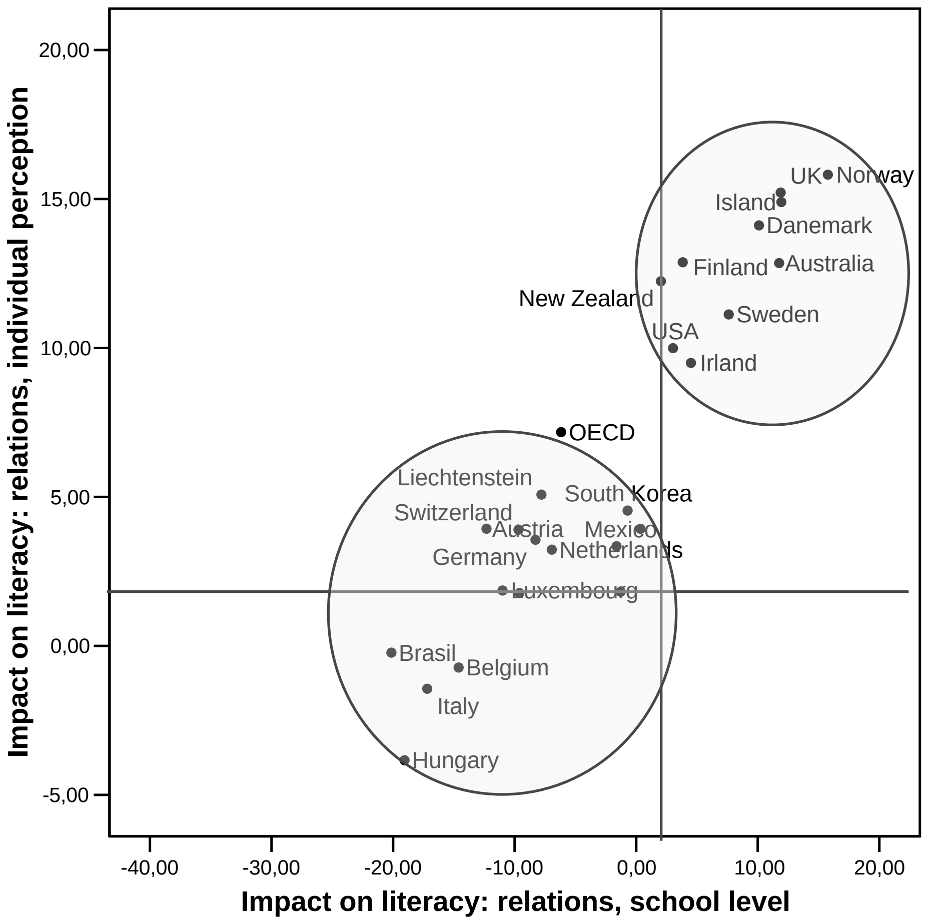
4. How LSA can inform EER

> exploring & generating hypotheses

> testing hypotheses

> understanding systemic and cultural effects

PISA 2000:: School- vs individual level effects of teacher-student-relations on reading literacy



5. Combining LSA and EER: longitudinal designs

> understanding school and classroom effects on growth in student motivation & achievement

> understanding school change over a period of several years

Two Paradigms (Reynolds 2005, p243)

School effectiveness	School improvement
Looking for generalizable results	Evaluation of school reform, „bottom up“
Focus: Outcomes	Focus: Organizational processes
Schools as static organizations	Schools as dynamic organizations
Quantitative	Quantitative
Longitudinal on the student level	**Longitudinal on the school level**

5.1 Longitudinal Design at the individual level: The German Study on Language Instruction (DESI)

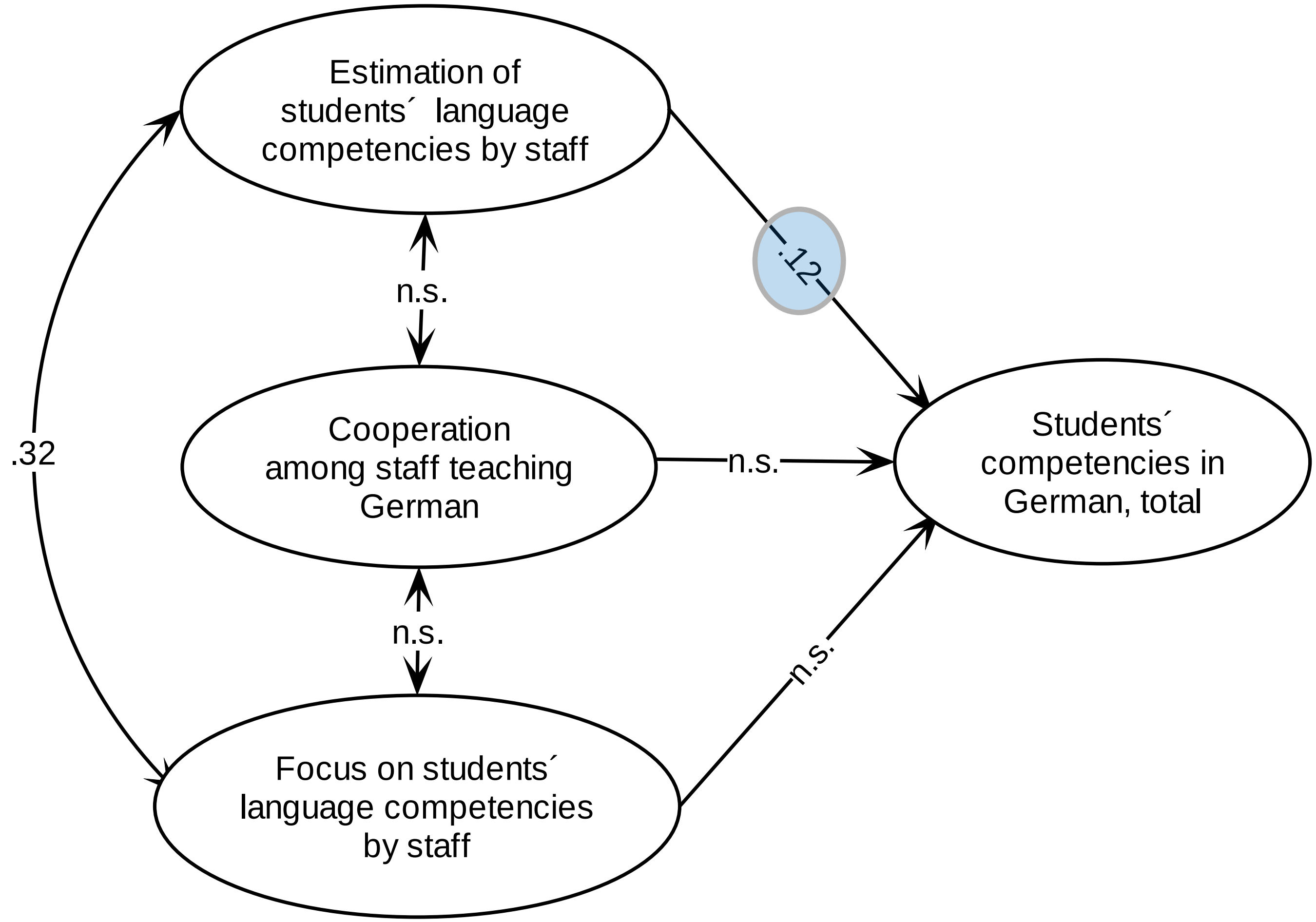
T1: 09-10/2003				**T2: 05-07/2004**
Language competencies in German & in English N = 10.543 *Achievement tests* Motivation & perception of instruction *Student questionnaire*	Individual and family background *Student questionnaire* *Parent questionnaire*	**School & instructional context of teaching and learning** *Principal questionnaire* ***Questionnaire for staff teaching German and English at school*** N = 210	Motivation and perception of instruction *Student questionnaire* *Teacher questionnaire (assessed classes)*	**Language competencies in German & in English** N = 10.543 *Achievement tests*
	11/2003 – 05/2004: Observation of English lessons by videotaping in 105 classrooms			

German Study on Language Instruction (Klieme et al., 2008)

Modeling individual, school and staff characteristics: 2-level SEM with Mplus

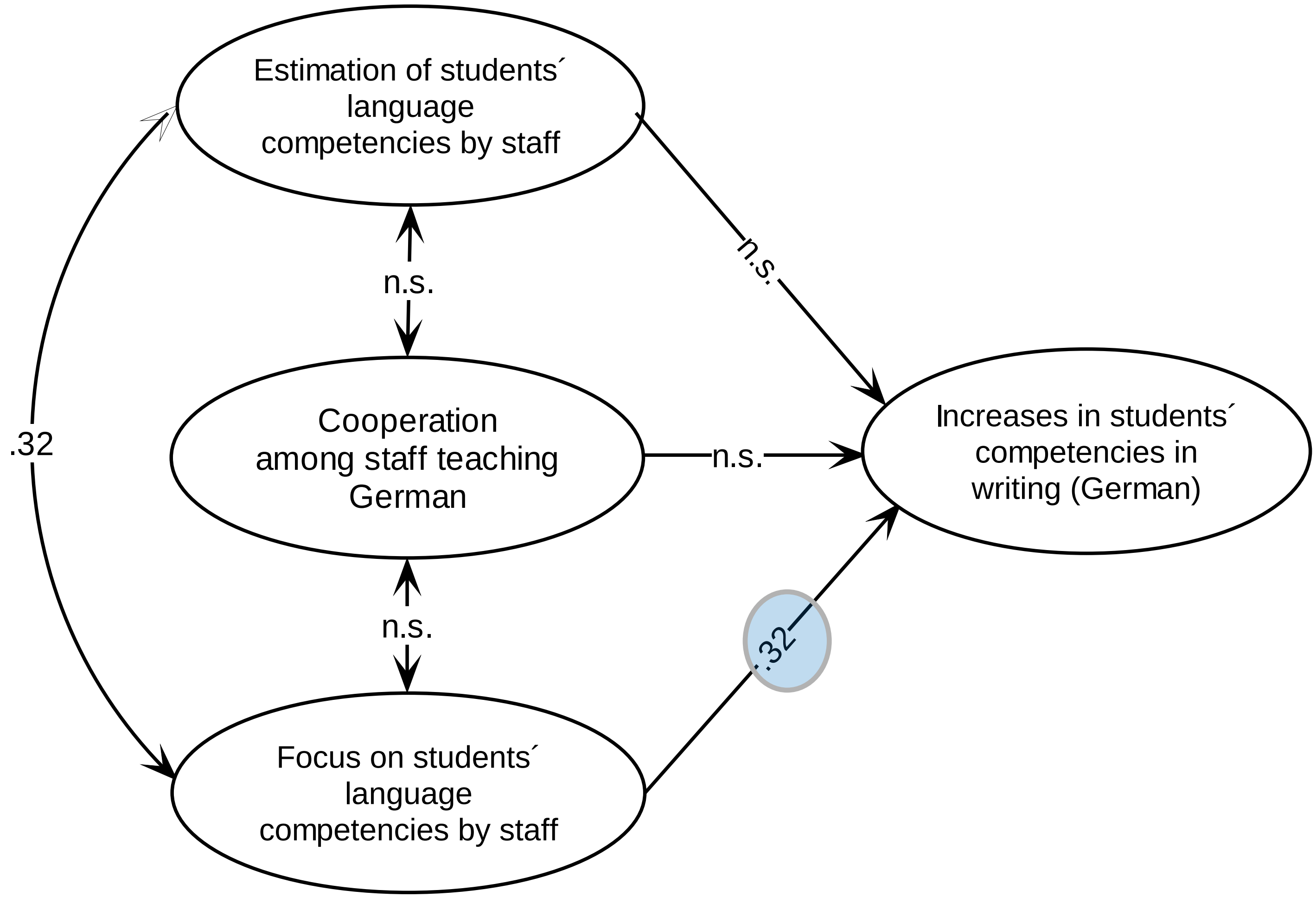
Background variables and **predictors**	**Criteria**
Individual level: • Gender • Cognitive ability test: KFT-N • SES: HISEI • German as first language • Attending bilingual lessons	*Yield in competencies: T2* • German achievement, total • English achievement, total *Increases in competencies: T1-T2* • German: writing/w.composition • English: C-Test
School level: • School types • cognitive, social and linguistic composition of students • School with bilingual classes • **Estimation of student achievement** • **Cooperative practices** • **Focus on language competencies**	

Relations between school processes (department staff), and student outcomes
(Steinert/Hartig/Klieme 2007; Klieme/Steinert/Hochweber 2010)



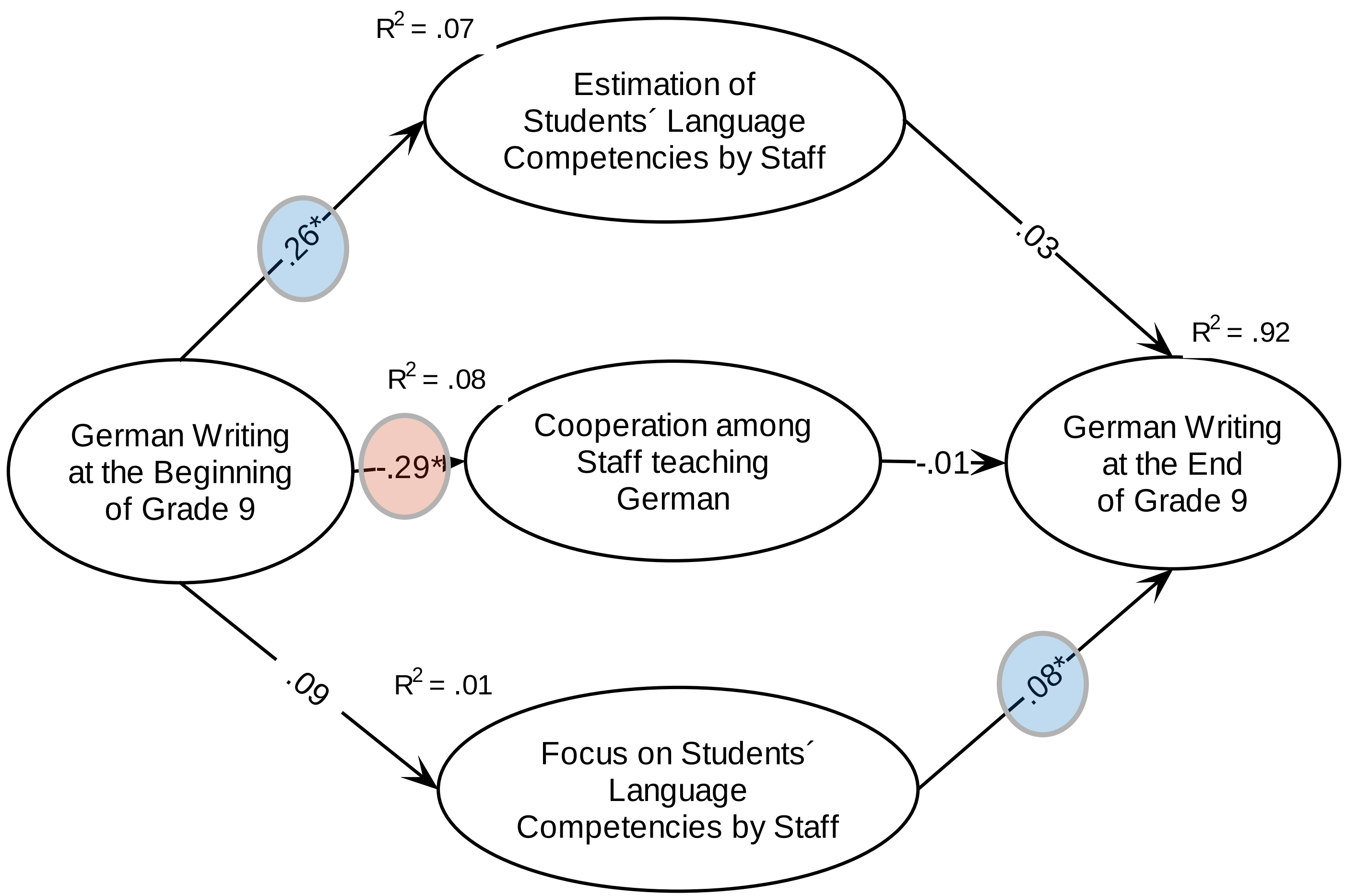
Modellfit: χ^2 = 158.713; *df* = 80; CFI = .97; RMSEA = .01; $SRMR_{between}$ = .08; $R^2_{between}$ = .86.

Relations between school processes (department staff), and student outcomes
(Steinert/Hartig/Klieme 2007; Klieme/Steinert/Hochweber 2010)



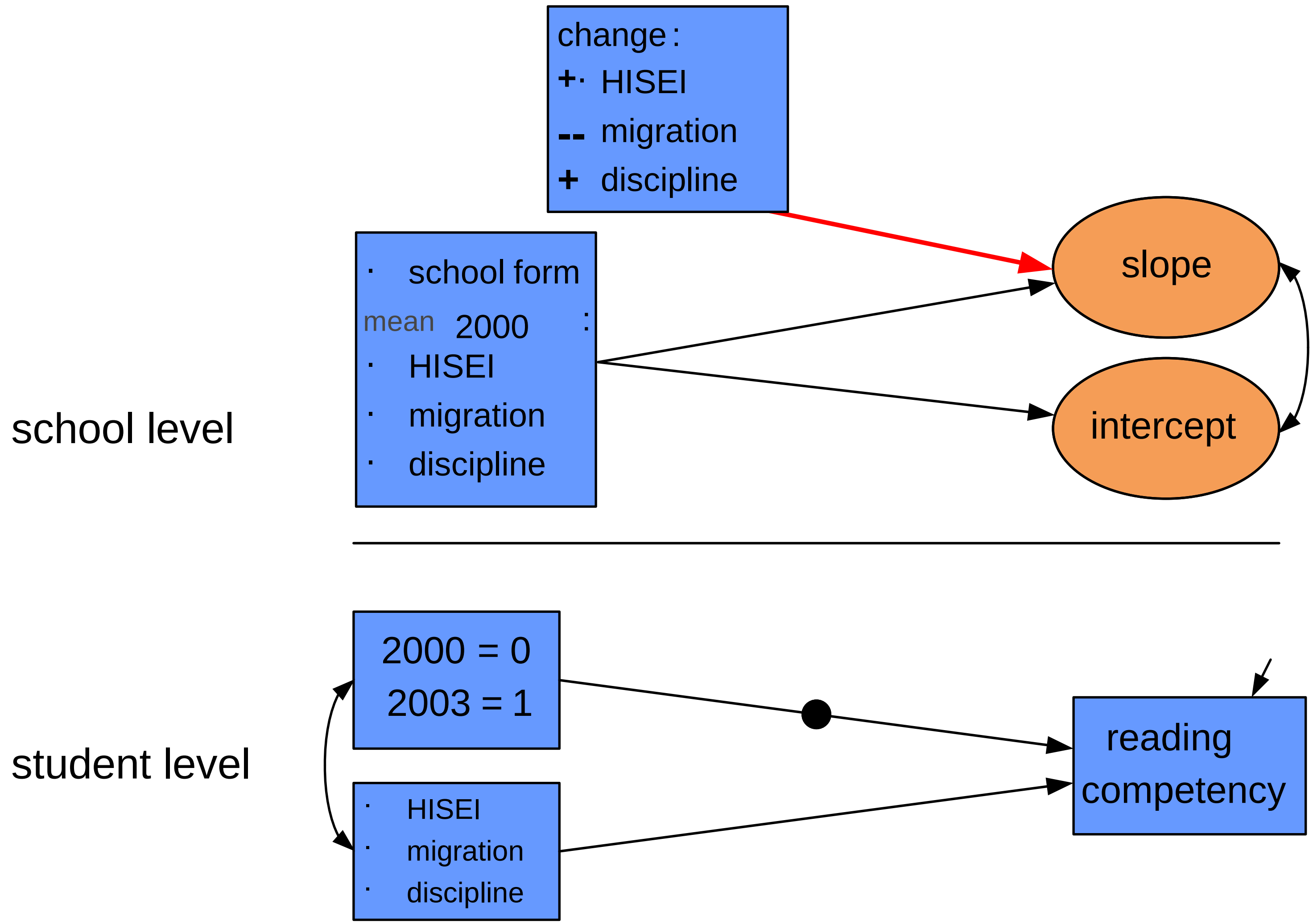
Modellfit: χ^2 = 151.044; *df* = 80; CFI = .92; RMSEA = .01; $SRMR_{between}$ = .07; $R^2_{between}$ = .21.

Relations between school processes (department staff), and student outcomes
(Steinert/Hartig/Klieme 2007; Klieme/Steinert/Hochweber 2010)



Modellfit: χ^2 = 641.061; *df* = 116; CFI = .92; RMSEA = .02; $SRMR_{between}$ = .16.

5.2 Longitudinal Design at the school level: The German PISA School Panel (n=369 schools) (Hochweber/Steinert/Gomolka/Klieme 2010)



6. Conclusions

- As a tool for policy making, LSA need not provide scientific explanations; descriptive validity is sufficient !
- However, in order to cover indicators that are relevant for professionals and policy makers, LSA should build on factors that have been shown elsewhere to have an impact on educational outcomes and change (!).
- (I)LSA can inform EER by generating and testing hypotheses, and especially by exploring system/culture effects.
- Causal inference requires either matching on the school level, or antecedent measures on the student level.
- Longitudinal designs (student/school panels) significantly enhance the power of LSA studies, but: BQs and tests have to be aligned with the analytical perspective
- There is still a lack of theoretical foundations for inferences on system effects and school change.