Outline

What this talk will cover

1. The Instrument of School Inspectorates in Germany

2. Concepts of Validity

3. Applying some Aspects of a Validity Framework to School Inspection

4. Summary & Recommendations

The Instrument of School Inspectorates in Germany

→ School inspectorates are expected to be the most powerful tool in advancing quality assurance, school improvement, innovation implementation, and school accountability (cf. Döbert & Dedering, 2008)

→ Between 2004 and 2008 all 16 states implemented some form of an external evaluation scheme for their schools

→ In Brandenburg, e.g., inspection of one school costs on average 10,000 EUR, summing up to 1.5 Mio EUR per year or 7.5 Mio EUR per inspection cycle for all public schools.

→ Legitimate question: Do school inspections yield valid assessments of school quality (improvement) that justify these expectations and costs?

Validity Concept

The concept of validity has historically seen a variety of iterations that involved "packing" different aspects into the concept and subsequently "unpacking" some of them.

Points of broad consensus

➡ Validity is the most fundamental consideration in the evaluation of the appropriateness of claims about, and uses and interpretations of assessment results.

➡ Validity is a matter of degree rather than all or none.

Validity Concept

Broad, but not universal agreement
(for a dissenting viewpoint, see Lissitz & Samuelson, 2007)

➡ It is the **uses** and **interpretations** of an assessment result, i.e. the **inferences**, rather than the assessment result itself that is validated.

➡ Validity may be relatively high for one use of assessment results by quite low for another use or interpretation.

Validity Concept

Main controversial aspect

➜ "The consequential aspect of (…) validity includes evidence and rationales for evaluating the intended and unintended consequences of score interpretation and use in both the short- and long-term" (Messick, 1995, p. 746)

➜ "[T]he validation process is scientific as well as rhetorical and requires both evidence and argument" (p. 747).

Samuel Messick´s Unitary Concept of Validity

→ Validation takes the form of an open-ended argument that evaluates the overall plausibility of the proposed interpretations of an assessment result and its consequences.

Construct validity					
Content	Substantive	Structural	Generalisability	External	Consequential
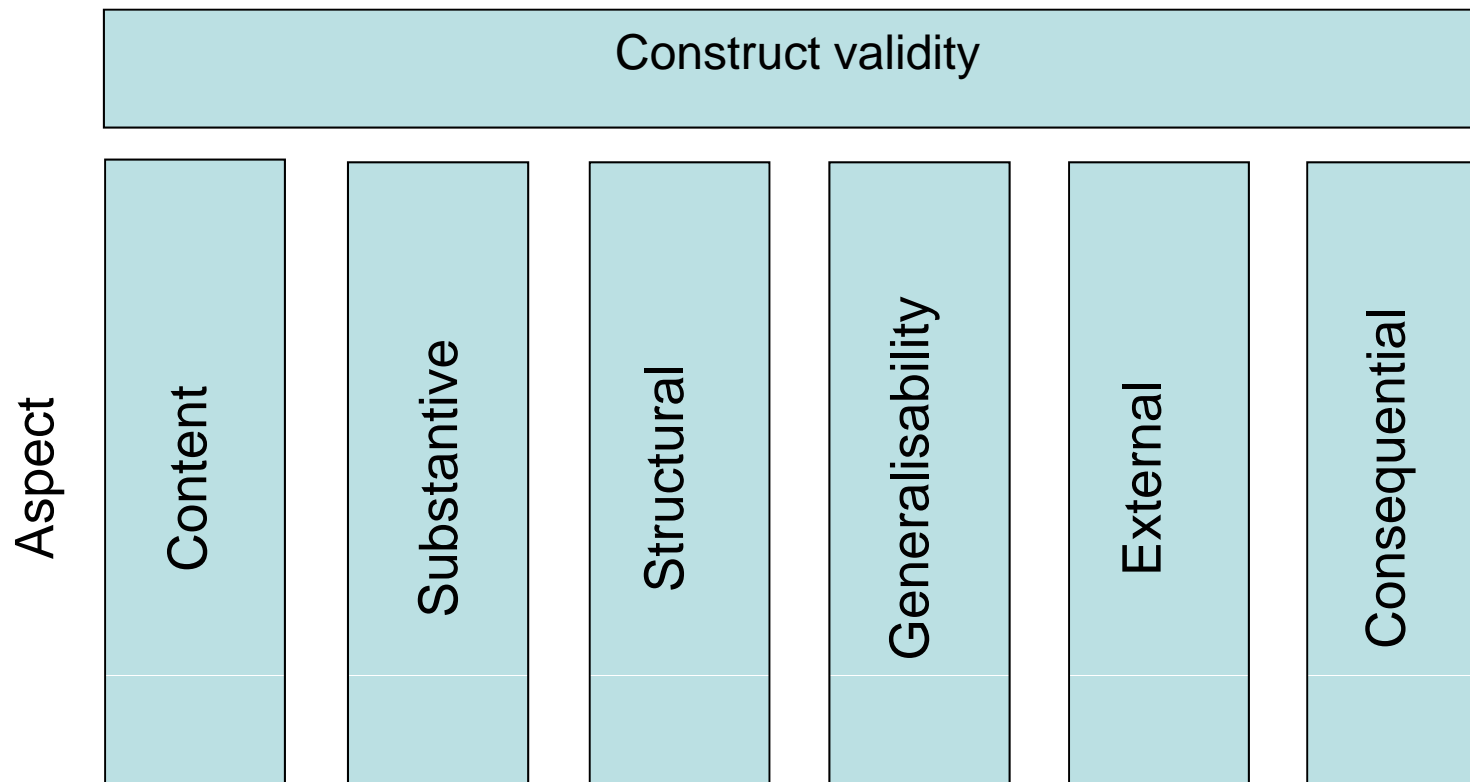
Aspect

Data Sources Used

[A]

➡ Secondary analysis of school inspection data from the state of Brandenburg.

➡ All schools inspected between school years 2005/06 and 2007/08 ($N = 368$).

[B]

➡ First wave of a survey of $N=173$ principals and $N=660$ teachers of schools inspected in the 2006/07 and 2007/08 academic years

➡ Survey covers the perceived effects and side-effects of past inspections

➡ Fundamentally, a school inspection amounts to a highly complex diagnosis of an organisation that draws on various methods from social research (observation, surveys, interviews, document analysis).

➡ Inspectors summarize and collate a considerable amount of information, so as to arrive at a final evaluation of the quality of a school.

➡ The school inspection in Brandenburg evaluates sixteen quality dimensions assigned to 6 "quality areas" (higher order dimensions) with a total of 78 indicator items.

Schulvisitation
im Land Brandenburg
Konzeption und Handbuch zur Schulvisitation

A typical school profile as a result of school inspection

Quality areas	Quality dimensions (QD)	Very strong	strong	weak	Very weak
OUTPUT	OD 1: Competencies				
	QD 2: School leaving certificates				
	QD 3: Satisfaction	X			
LEARNING CULTURE	QD 4: Internal school curricula	X			
	QD 5: Classroom management	X			
	QD 6: Activation of learning processes			X	
	QD 7: Goal-oriented learning processes		X		
	QD 8: Class and working atmosphere		X		
	QD 9: Transparent achievement expectations		X		
	QD 10: Diagnostic and individual promotion		X		
SCHOOL CULTURE	QD 11: Active participation	X			
	QD 12: Co-operation		X		
	QD 13: School leadership		X		
SCHOOL MANAGEMENT	QD 14: Quality management			X	
	QD 15: Organization	X			
PROFESSIONAL DEVELOPM.	QD 16: Professional development		X		
QUALITY DEVELOPMENT	QD 17: School program		X		
	QD 18: School self-evaluation			X	

A typical example of a quality dimension in the evaluation process in Brandenburg school inspection

Quality Dimension 4: Internal School Curricula		Rating			
Source	Indicator	+	o	−	#
DA, HS, HD, T, TQ	4.1 Internal school curricula are developed and coordinated in school conferences	☒	☐	☐	☐
DA, HS, HD, T	4.2 Internal school curricula describe competence domains	☐	☒	☐	☐
DA, HS, HD, T, TQ	4.3 Interdisciplinary elements are components of school instruction planning	☒	☐	☐	☐
HS, HD, T	4.4 Teachers co-ordinate methodological and subject-related questions among themselves	☒	☐	☐	☐
DA, HS, HD, T, S, P, PvE	4.5 The goals and content of instruction are transparent to students, parents as well as partners for vocational education	☐	☒	☐	☐
DA, HS, HD, S, T, TQ	4.6 Learning opportunities presented by the school include aspects of practical learning	☐	☐	☒	☐
HS, HD, T, S	4.7 School internal curricula contain specifications for the application of media	☒	☐	☐	☐

Requirement for meeting standard: The standard for this dimension is met when Indicator 4.1 is positive together with three of the other indicators 4.2 to 4.7. Total Score: 4 ☐ 3 ☒ 2 ☐ 1 ☐

Note. HS = interview with head of school; HD = interview with head of school department; T = teacher interview; S = student interview; P = parent interview; PvE = interview with partners of vocational education; OB = lesson observation; DA = data collection questionnaire; TQ = teacher questionnaire; "+" = positive contribution; "o" = little contribution; "−"= negative contribution; # = no rating possible; 1 = mostly weak (no indicator "+"); 2 = more weak than strong (standard not fulfilled); 3 = more strong than weak (standard fulfilled); 4 = mostly strong (85 % of all indicators "+" and none "−").

The Claims of School Inspection in Brandenburg

Claims of Brandenburg's inspectorate:

➡ Evaluation of a school's performance by means of assessing process quality and output quality

➡ Increasing the validity of a school's self-evaluation through an external, more objective view

➡ Advancing the within-school processes of discussion and development

➡ Monitoring school performance for evidence-informed educational policy decisions

Applying Messick's validity framework

Aspect	Content	Substantive	Structural	Generalisability	External	Consequential
Construct validity						
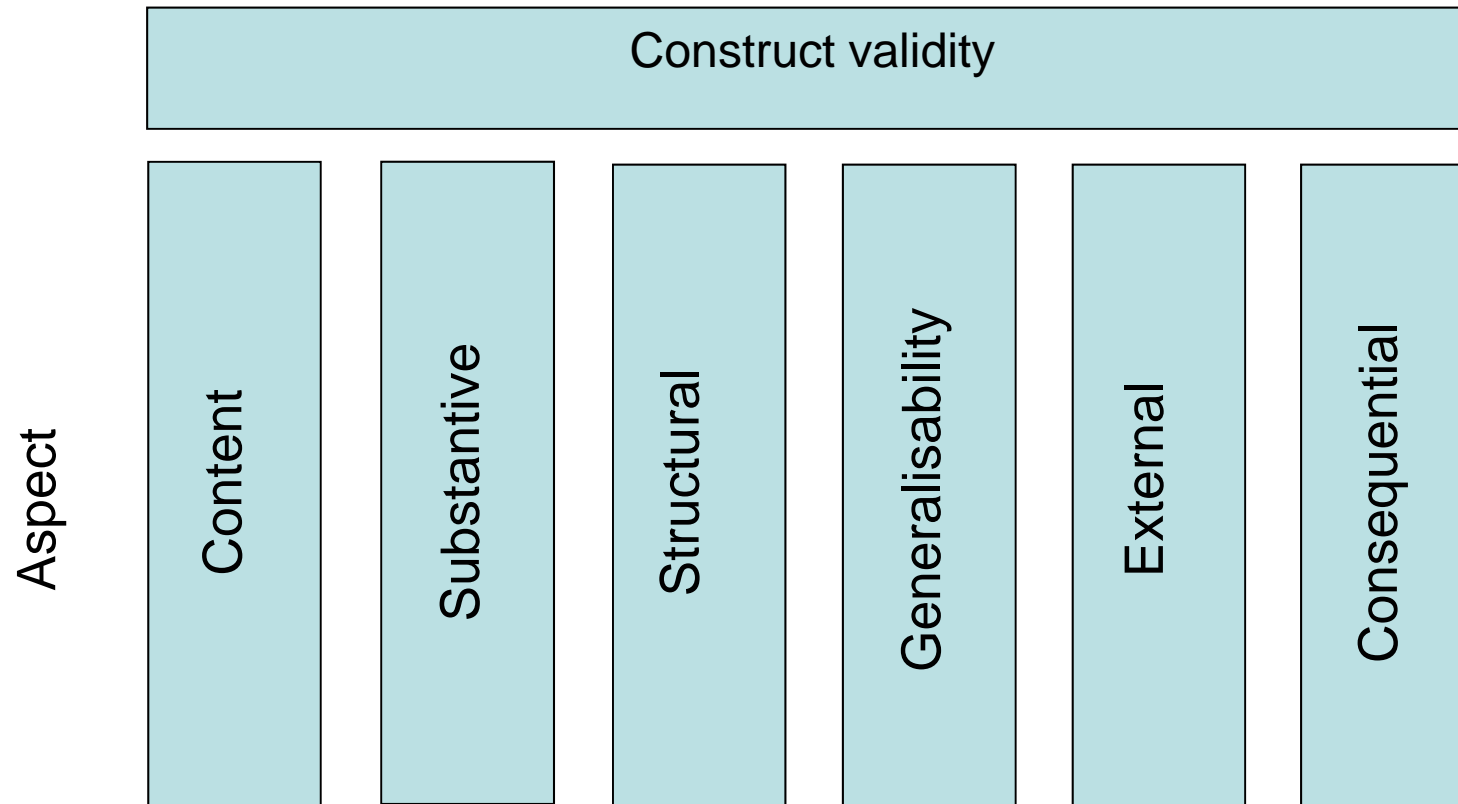
Content aspect
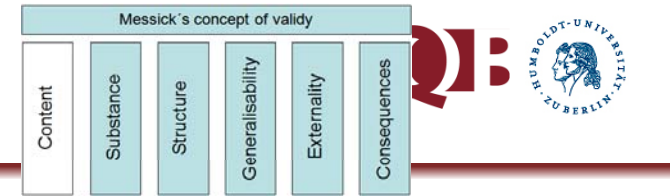
Content validity requires evidence of
→ the relevance of measured (test) content
→ of its representativeness for the construct
→ Sources of potential invalidity are under-representation of the construct and construct-irrelevant variance

Content validity in the context of school inspection
→ Does school inspection assess the relevant and representative aspects?
→ Do assessment instruments reflect the conceptual framework of reference?

Open questions:
→ Is the sample of a school's reality as assessed by school inspection relevant and representative for school quality and instructional quality?
→ On what rational basis can standards of good school and instructional quality be established?

Potential validation strategies
→ Content validation of quality frameworks (Scheerens et al., 2005)
→ Establishing standards in formal standard setting procedures

Substantive aspect
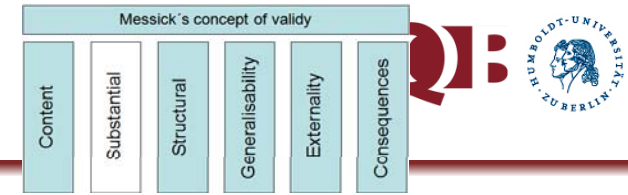
Substantial validity in the context of school inspection
➜ Are the relevant processes assumed to affect school quality adequately represented in school inspection?

On the classroom level
➜ Are classroom observations suitable for capturing the instructional processes influencing students' learning?

On the school level
➜ Which attributes at the organizational level (such as school management, organization of instruction, school self-evaluation) contribute to improvements in instructional quality?

Open questions:
➜ Is the school level at all the adequate aggregate level to reflect instructional quality?

Potential validation strategies
➜ Multi-level approaches to integrate classroom level with school level processes (cross-level effects)

Structural aspect
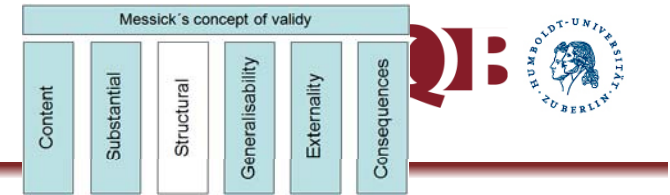
Does the measurement model conform with the structures of the construct?

Structural validity in the context of school inspection
➜ Assignment of indicators of school quality to quality dimensions as described in the quality framework should be reflected empirically in the relationships between actually assessed variables.

Open questions:
➜ Is each of the computational and aggregational steps made during an inspection evaluation sound?

Potential validation strategies
➜ (Confirmatory) dimensional analysis
➜ Reliability analysis

Structural aspect:
Is it sound to combine quality ratings into an overall score?

Quality Dimension	N	n_{item}	α
Transparent performance requirements	362	4	.19
School co-operation	368	4	.19
Instructional organisation	363	3	.27
Internal school curricula	338	7	.41
Classroom management	368	3	.45
Class and work environment	365	5	.45
Diagnostics and individual support	365	4	.46
Professional development	327	5	.47
Satisfaction	357	6	.49
Active involvement in school life	363	5	.51
Quality improvement through evaluation	362	5	.55
Active learning process	368	4	.58
Goal-oriented learning process	366	6	.58
School development plan	229	4	.60
Quality management	360	6	.72
Management responsibility	363	5	.77

Average Cronbach's α=0.51

Note. Low intercorrelations can also result from the sometimes extremely low variance in individual indicator evaluations

Structural aspect
Exploration of the dimensionality of inspection results

Factor 1	Factor loadings	Dimension as defined in quality framework
School development plan	.73	6: Goals and strategies of quality improvement
Quality improvement through evaluation	.69	6: Goals and strategies of quality improvement
Internal school curricula	.59	2: Teaching and learning – Instruction
Professional development	.48	5: Professionality of teaching staff
Quality management	.46	4: School management
Transparent performance requirements	.41	2: Teaching and learning – Instruction
Factor 2		
Class and work environment	.80	2: Teaching and learning – Instruction
Classroom management	.79	2: Teaching and learning – Instruction
Active learning process	.64	2: Teaching and learning – Instruction
Factor 3		
Satisfaction	.76	1: Results of the school
Instructional organisation	.72	4: School management
Management responsibility	.61	4: School management
Diagnostics and individual support	.58	2: Teaching and learning – Instruction
Active involvement in school life	.58	3: School culture
Factor 4		
Goal-oriented learning process	.86	2: Teaching and learning – Instruction
Factor 5		
School co-operation	.81	3: School culture

Note. A principal component analysis with subsequent oblique rotation yielded 5 factors with an eigenvalue of >1 and an explained variance of 57% overall.

Generalisability aspect
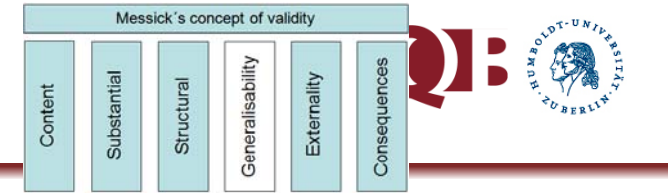
Generalisability
  ➡ The extend to which research findings can be applied to settings other than those in which they were originally applied.
  ➡ Relates to generalisations of test results across different points in time, situations and evaluators.

Generalisability in the context of school inspection
  ➡ Are the lessons chosen representative for the instructional "culture"?
  ➡ To what extent depend results of inspections on the inspectors themselves?
  ➡ How open to manipulation are the results of school inspection?

Open questions:
  ➡ What are the main construct-irrelevant context factors that might threaten the generalisability and therefore validity of the inspection results?

Potential validation strategies
  ➡ Interrater reliability of inspectors (de Jong & Reezigt, 2007)
  ➡ Multifaceted Rasch models (Pietsch & Tosana, 2008)

External aspect
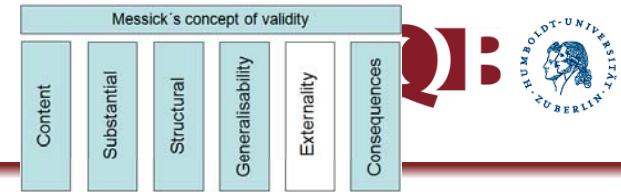
External validity concerns the relationships between test results and external criteria.

External validity in the context of school inspection
- ➜ Does the proficiency of students develop more favourably in a school with strong process qualities as evaluated by the school inspection than in a comparable school with weaker attested process qualities?

Open questions:
- ➜ What are the expected effects (effect size, time lag of effects) of inspection results feedback on a schools performance?
- ➜ What kind of performance data are needed to evaluate the effects of school inspection

Potential validation strategies
- ➜ Linking the results of school inspections to longitudinal data on students' performance trajectories
- ➜ As an approximation: Linking cross-sectional data of state-wide assessments to school inspection data across years.

Consequential aspect
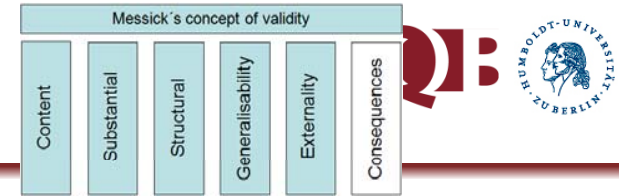
The aspect of *consequential validity* relates to whether the diagnosis has had the desired effects and whether there have been any unintended effects.

Consequential validity in the context of school inspection
- ➡ Does the inspectors' feedback report actually lead to school and instructional improvement?
- ➡ Are there unintended side-effects?

Open questions: Extant research is scarce and inconclusive
- ➡ Matthews & Sammons (2004): Study conducted by Ofsted - school inspections do improve school quality
- ➡ Cuckle & Broadhead (1999): undesirable side-effects of inspections, such as extreme pressure on staff and school management in the lead-up

Potential validation strategies
- ➡ Comparing the organisational development in inspected vs. not-inspected schools

Berlin-Brandenburg Study

➜ Theory to explain the different effects of school inspection of Ehren & Visscher (2006):

Feedback should be given differently depending on the type of school :

➜ for *low innovation capacity schools* it should include instructions for action.

➜ for *high innovation capacity schools* only strengths and weaknesses should be reported to promote an independent development.

➜ Research questions::

Feedback should be given differently depending on the type of school :

➜ How do school principals and teachers (department leaders) perceive school inspections and feedback from inspection reports?

➜ Can we identify types of reaction patterns to inspection?

→ Data base: school principals teachers from *N*=391 schools that were inspected in school years 2006/07 und 2007/08

Response rate	Berlin		Brandenburg		**Overall**	
	absolute	in%	absolute	in%	absolute	in%
School principals	143 of 283	50,5	173 of 278	62,2	316 of 561	**56,3**
Teachers	412 of 1132	36,4	660 of 986	66,9	1072 of 2118	**50,6**

→ Non-responder analysis yielded no evidence for a biased of self-selection process with respect to school level achievement data and overall inspection result.

→ Schooltypes of feedback use from inspection were identified by Latent Class Analysis (LCA).

Theoretical background

Inspectorate as a feedback system with a focus on school processes

Descriptive model based on Helmke & Hosenfeld (2005), Ehren & Visscher (2006) and Sommer (2009)
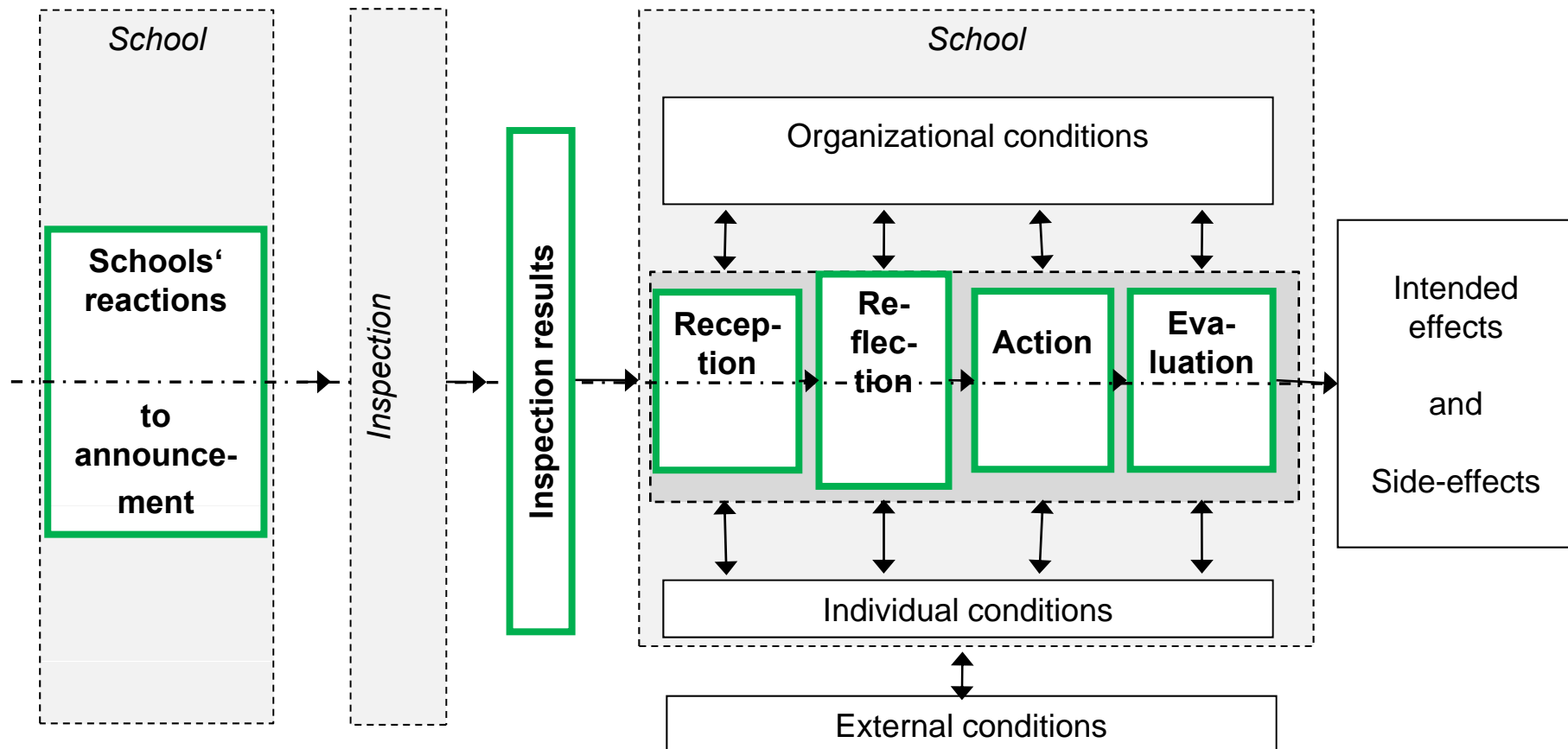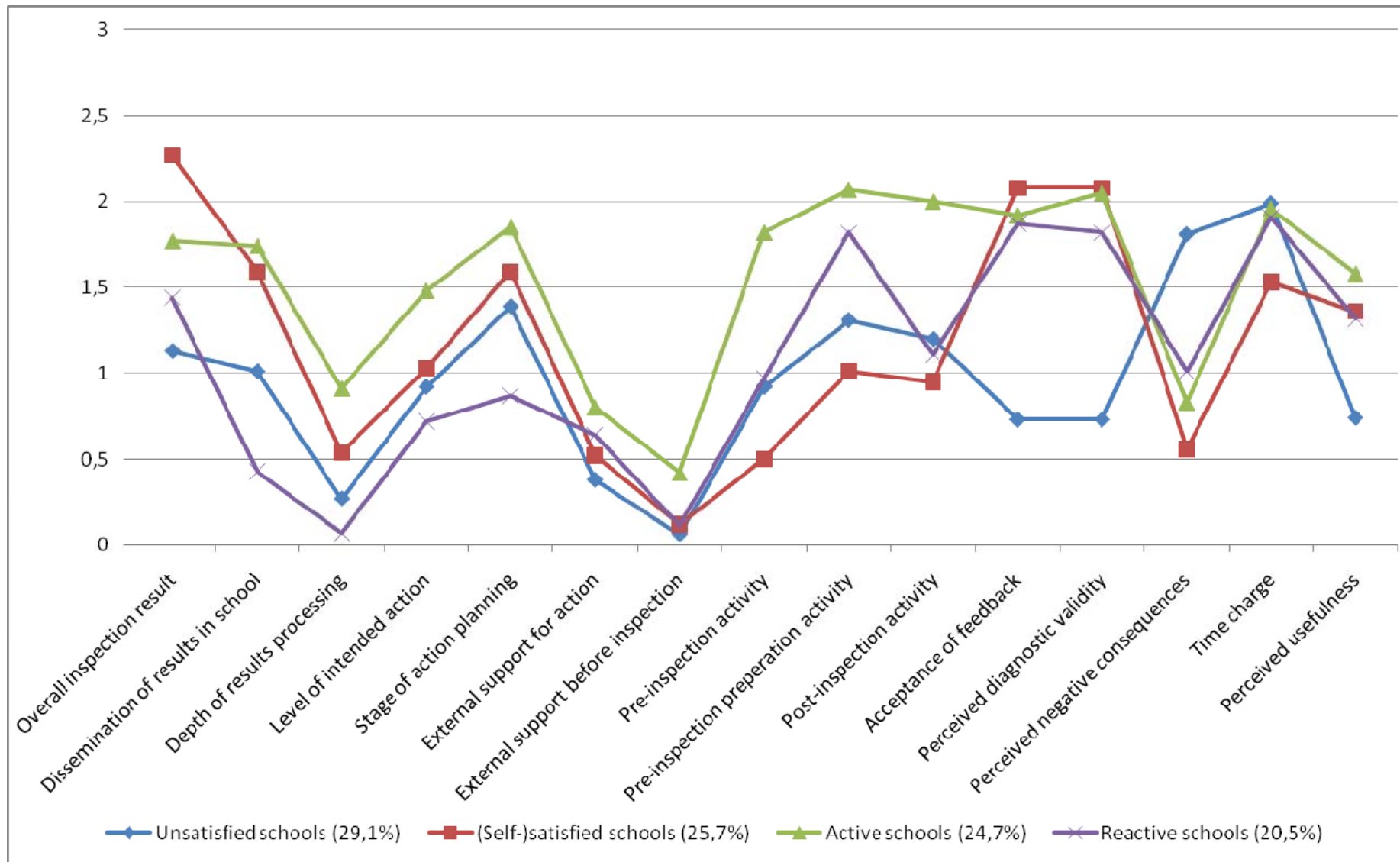
Schooltyp: „(Self-)satisfied" Schools (26%)
➡ Best overall inspection result
➡ Low level of activity
➡ Comprehensive dissemination and results processing
➡ Overall positive perception on inspection process (acceptance, benefit, diagnostic validity, low perceived time charge)

Legend:
- Unsatisfied schools (29,1%)
- (Self-)satisfied schools (25,7%)
- Active schools (24,7%)
- Reactive schools (20,5%)

X-axis labels: Overall inspection result; Dissemination of results in school; Depth of results processing; Level of intended action; Stage of action planning; External support for action; External support before inspection; Pre-inspection activity; Pre-inspection preperation activity; Post-inspection activity; Acceptance of feedback; Perceived diagnostic validity; Perceived negative consequences; Time charge; Perceived usefulness

Schooltype: „Active" Schools (25%)
� Good overall inspection result
➡ High self-reported activity level
➡ Highest level of dissemination and results processing
➡ Highest level of demand of support
➡ High acceptance of inspection procedure despite time stress

Legend:
— Unsatisfied schools (29,1%) — (Self-)satisfied schools (25,7%) — Active schools (24,7%) — Reactive schools (20,5%)

X-axis categories: Overall inspection result; Dissemination of results in school; Depth of results processing; Level of intended action; Stage of action planning; External support for action; External support before inspection; Pre-inspection activity; Pre-inspection preperation activity; Post-inspection activity; Acceptance of feedback; Perceived diagnostic validity; Perceived negative consequences; Time charge; Perceived usefulness

Results of LCA: 4 classes

Schooltype: „Reactive" Schools (21%)
- → High on pre-inspection preperation activity
- → Low on school-related activities, as well as lowest dissemination of results, results processing and contemplation
- → Positive estimation of inspection feedback

Legend:
- Unsatisfied schools (29,1%)
- (Self-)satisfied schools (25,7%)
- Active schools (24,7%)
- Reactive schools (20,5%)

X-axis categories: Overall inspection result, Dissemination of results in school, Depth of results processing, Level of intended action, Stage of action planning, External support for action, External support before inspection, Pre-inspection activity, Pre-inspection preperation activity, Post-inspection activity, Acceptance of feedback, Perceived diagnostic validity, Perceived negative consequences, Time charge, Perceived usefulness

Schooltype: „Unsatisfied" Schools (29%)
→ Worst overall result
→ Average dissemination and results processing
→ Low acceptance of results and perceived validity
→ High time charge for little perceived use

Legend: Unsatisfied schools (29,1%) — (Self-)satisfied schools (25,7%) — Active schools (24,7%) — Reactive schools (20,5%)
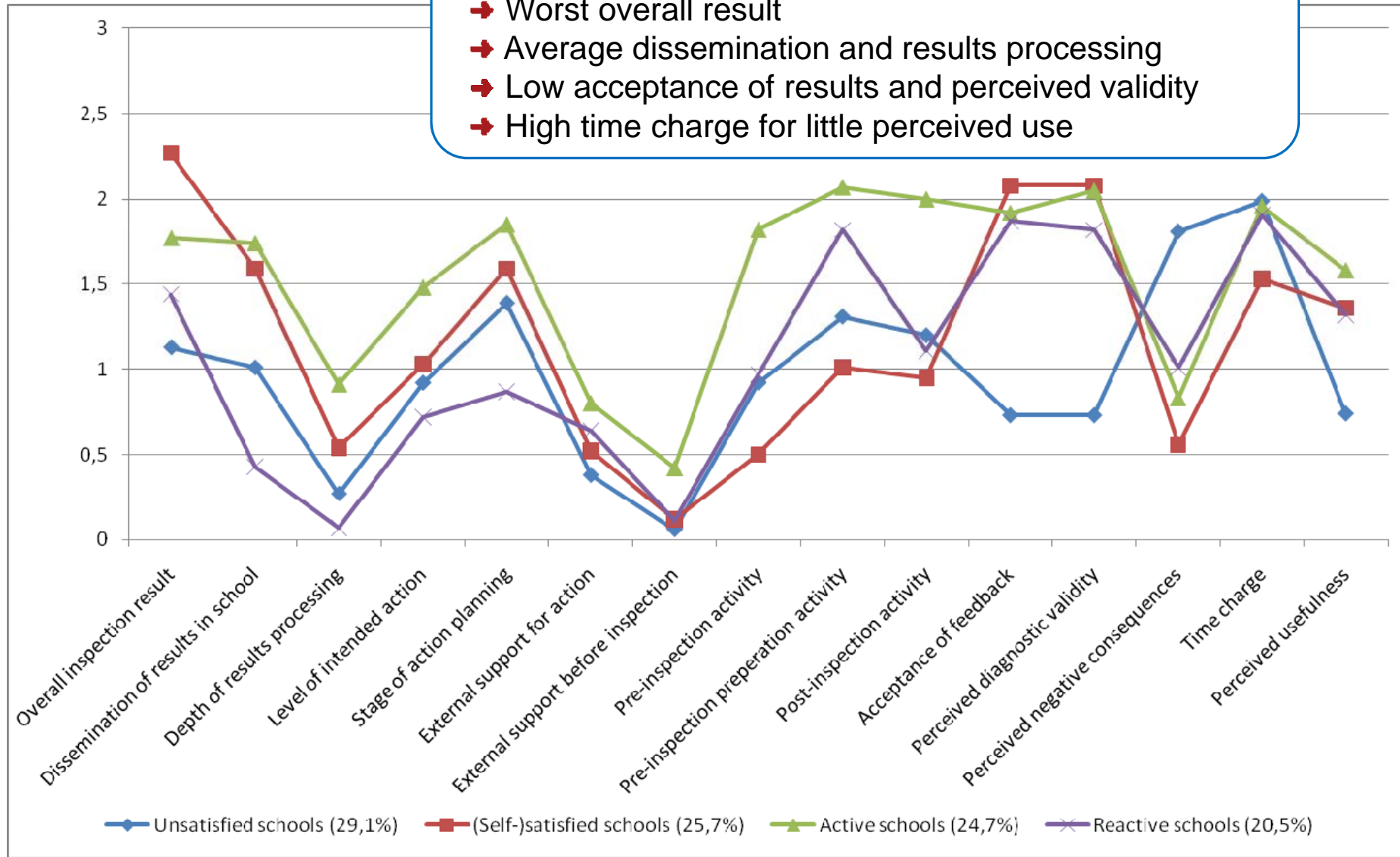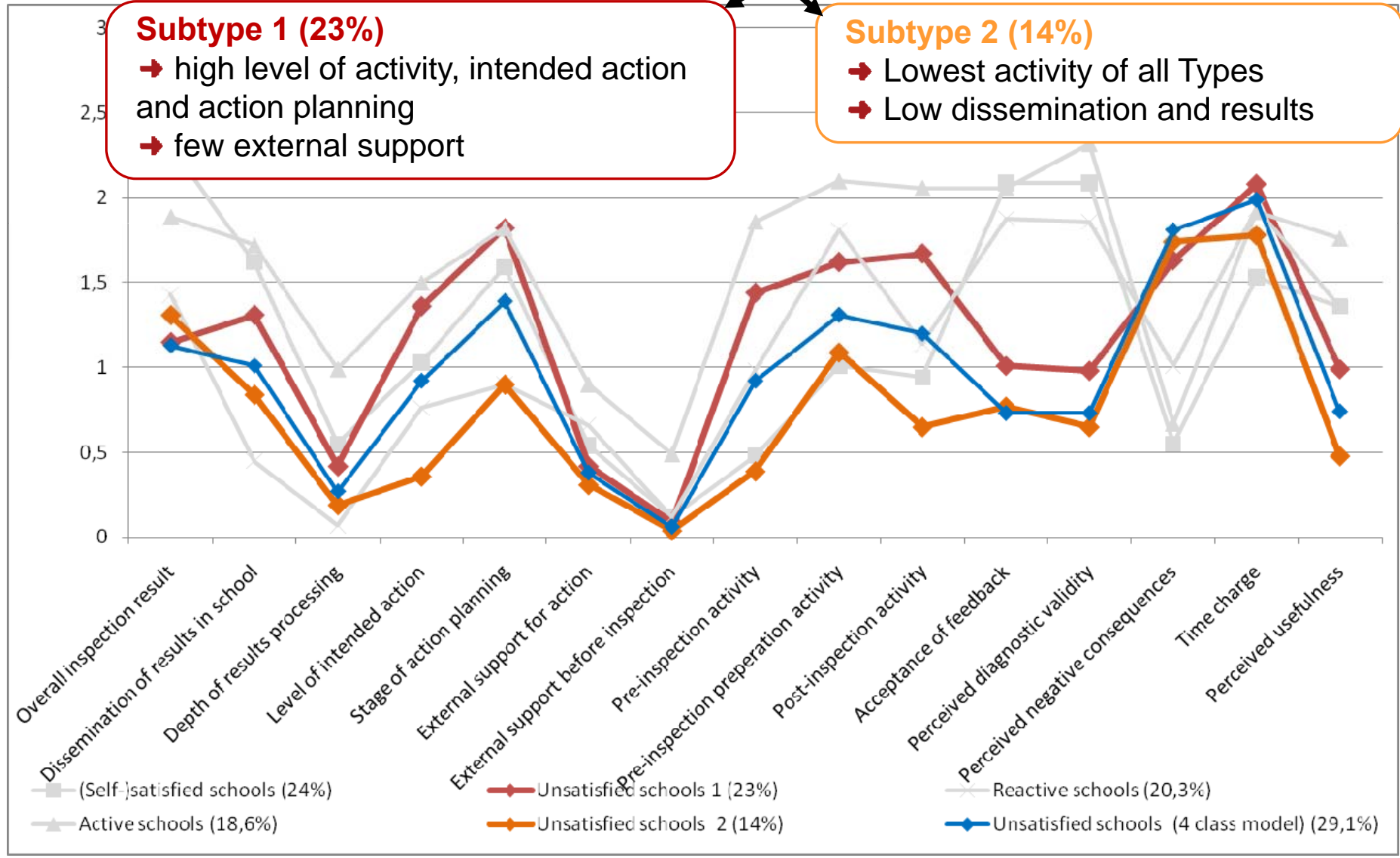
Schooltype: „Unsatisfied" Schools
➜ Worst overall result
➜ Negative perception of inspection process

Subtype 1 (23%)
➜ high level of activity, intended action and action planning
➜ few external support

Subtype 2 (14%)
➜ Lowest activity of all Types
➜ Low dissemination and results

Legend:
- (Self-)satisfied schools (24%)
- Active schools (18,6%)
- Unsatisfied schools 1 (23%)
- Unsatisfied schools 2 (14%)
- Reactive schools (20,3%)
- Unsatisfied schools (4 class model) (29,1%)

X-axis categories:
Overall inspection result, Dissemination of results in school, Depth of results processing, Level of intended action, Stage of action planning, External support for action, External support before inspection, Pre-inspection activity, Pre-inspection preperation activity, Post-inspection activity, Acceptance of feedback, Perceived diagnostic validity, Perceived negative consequences, Time charge, Perceived usefulness

Summary & Recommendations

→ School inspection is implemented in all federal states of Germany and many other countries.

→ Results of school inspection are supposed to drive processes of school improvement.

→ However, applying Messick's concept of validity to school inspection uncovers several unresolved questions about the soundness of the inferences drawn from inspection results.

→ Future cooperation between educational administration staff and researchers should strive to reconcile scientific rigor in quantitative assessment with the sensitivity of experienced inspection personnel.

→ Emphasis should be placed on measurement models first, followed by evaluation of structural hypotheses to validate school inspection results

→ From a strategic point of view consequential aspects of validity must be addressed with priority in order to gain acceptance in schools.

Thank you for your attention!

SICI Conference 2010

North Rhine-Westphalia
„Quality Assurance in the Work of Inspectors"