How valid are school inspections?
Problems and strategies for validating processes and results

Hans Anand Pant

SICI Conference 2010
North Rhine-Westphalia
„Quality Assurance in the Work of Inspectors“
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
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?
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.
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).
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.
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.
A typical school profile as a result of school inspection

<table>
<thead>
<tr>
<th>Quality areas</th>
<th>Quality dimensions (QD)</th>
<th>Very strong</th>
<th>strong</th>
<th>weak</th>
<th>Very weak</th>
</tr>
</thead>
<tbody>
<tr>
<td>OUTPUT</td>
<td>OD 1: Competencies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>QD 2: School leaving certificates</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>QD 3: Satisfaction</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LEARNING CULTURE</td>
<td>QD 4: Internal school curricula</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>QD 5: Classroom management</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>QD 6: Activation of learning processes</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>QD 7: Goal-oriented learning processes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>QD 8: Class and working atmosphere</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>QD 9: Transparent achievement expectations</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>QD 10: Diagnostic and individual promotion</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCHOOL CULTURE</td>
<td>QD 11: Active participation</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>QD 12: Co-operation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>QD 13: School leadership</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCHOOL MANAGEMENT</td>
<td>QD 14: Quality management</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>QD 15: Organization</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PROFESSIONAL DEVELOPM.</td>
<td>QD 16: Professional development</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QUALITY DEVELOPMENT</td>
<td>QD 17: School program</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>QD 18: School self-evaluation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A typical example of a quality dimension in the evaluation process in Brandenburg school inspection

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

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 “–”).
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

Construct validity

- Content
- Substantive
- Structural
- Generalisability
- External
- Consequential
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?

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

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**

<table>
<thead>
<tr>
<th>Factor 1</th>
<th>Factor loadings</th>
<th>Dimension as defined in quality framework</th>
</tr>
</thead>
<tbody>
<tr>
<td>School development plan</td>
<td>.73</td>
<td>6: Goals and strategies of quality improvement</td>
</tr>
<tr>
<td>Quality improvement through evaluation</td>
<td>.69</td>
<td>6: Goals and strategies of quality improvement</td>
</tr>
<tr>
<td>Internal school curricula</td>
<td>.59</td>
<td>2: Teaching and learning – Instruction</td>
</tr>
<tr>
<td>Professional development</td>
<td>.48</td>
<td>5: Professionality of teaching staff</td>
</tr>
<tr>
<td>Quality management</td>
<td>.46</td>
<td>4: School management</td>
</tr>
<tr>
<td>Transparent performance requirements</td>
<td>.41</td>
<td>2: Teaching and learning – Instruction</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Factor 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Class and work environment</td>
<td>.80</td>
</tr>
<tr>
<td>Classroom management</td>
<td>.79</td>
</tr>
<tr>
<td>Active learning process</td>
<td>.64</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Factor 3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Satisfaction</td>
<td>.76</td>
</tr>
<tr>
<td>Instructional organisation</td>
<td>.72</td>
</tr>
<tr>
<td>Management responsibility</td>
<td>.61</td>
</tr>
<tr>
<td>Diagnostics and individual support</td>
<td>.58</td>
</tr>
<tr>
<td>Active involvement in school life</td>
<td>.58</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Factor 4</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal-oriented learning process</td>
<td>.86</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Factor 5</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>School co-operation</td>
<td>.81</td>
</tr>
</tbody>
</table>

*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 school's 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
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?
Participants and Methods

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

<table>
<thead>
<tr>
<th></th>
<th>Berlin</th>
<th>in%</th>
<th>Brandenburg</th>
<th>in%</th>
<th>Overall</th>
<th>in%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Response rate</strong></td>
<td>absolute</td>
<td></td>
<td>absolute</td>
<td></td>
<td>absolute</td>
<td></td>
</tr>
<tr>
<td>School principals</td>
<td>143 of 283</td>
<td>50,5</td>
<td>173 of 278</td>
<td>62,2</td>
<td>316 of 561</td>
<td>56,3</td>
</tr>
<tr>
<td>Teachers</td>
<td>412 of 1132</td>
<td>36,4</td>
<td>660 of 986</td>
<td>66,9</td>
<td>1072 of 2118</td>
<td>50,6</td>
</tr>
</tbody>
</table>

Non-responder analysis yielded no evidence for a biased 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)
Results of LCA: 4 classes
Results of LCA: 4 classes

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)
Results of LCA: 4 classes

- 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
Results of LCA: 4 classes

Schooltype: „Reactive” Schools (21%)
- High on pre-inspection preparation activity
- Low on school-related activities, as well as lowest dissemination of results, results processing and contemplation
- Positive estimation of inspection feedback
Results of LCA: 4 classes

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
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
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!