Student Assessment in Germany: Present and Future Initiatives
Prof. Petra Stanat, Ph.D.

Presentation at the National Assessment Governing Board Meeting
Washington, DC
November 17, 2017
The so-called „PISA shock“

Results of PISA 2000 for Germany

- mean reading literacy significantly below the mean of OECD member states
- large variance of achievement scores
- particularly poor results at the lower end of the achievement distribution
- pronounced disparities associated with students’
  - socio-economic background
  - migration background
- large achievement differences among the 16 states
Comprehensive strategy for educational monitoring of the 16 states

1. Participation in international large-scale assessments of student achievement (PISA, PIRLS, TIMSS)

2. Testing and implementation of educational standards for primary school, secondary level I, and secondary level II
   • national assessment studies at the system level in primary schools (grade 4) and secondary level I (grade 9)
   • pool of tasks for school-leaving exam qualifying for university admission (grade 12/13)

3. Tools for quality assurance at the school level (e.g., comparison tests VERA in grades 3 and 8)

4. Comprehensive educational reporting (every 2 years)
Comprehensive strategy for educational monitoring of the 16 states

1. Participation in international large-scale assessments of student achievement (PISA, PIRLS, TIMSS)

2. Testing and implementation of educational standards for primary school, secondary level I, and secondary level II
   • national assessment studies at the system level in primary schools (grade 4) and secondary level I (grade 9)
   • pool of tasks for school-leaving exam qualifying for university admission (grade 12/13)

3. Tools for quality assurance at the school level (e.g., comparison tests VERA in grades 3 and 8)

4. Comprehensive educational reporting (every 2 years)
Comprehensive strategy for educational monitoring of the 16 states

1. Participation in international large-scale assessments of student achievement (PISA, PIRLS, TIMSS)

2. Testing and implementation of educational standards for primary school, secondary level I, and secondary level II
   - national assessment studies
   - leaving exam qualifying for university entrance (grade 12/13)

3. Tools for quality assurance at the school level (e.g., comparison tests VERA in grades 3 and 8)

4. Comprehensive educational reporting (every 2 years)
The Institute for Educational Quality Improvement

• founded in 2004
• independent academic institute, located at the Humboldt University Berlin
• financed by the 16 federal states („Länder“) in Germany
• interdisciplinary team psychologists, educational researchers, teachers, psychometricians
• expertise on subject-matter content: cooperation with other universities and research institutes
Basis of the national assessment system: National Educational Standards

- Adopted for core subjects by the Standing Conference of the Ministers of Education and Cultural Affairs (KMK)
- Describe core elements of knowledge and skills students should, on average, have acquired by the end of a certain stage in their educational career
- Implementation is mandatory for all 16 states
- Alignment of state-specific curricula and central exams with the national standards
## Educational Standards in Germany

<table>
<thead>
<tr>
<th></th>
<th>Primary school</th>
<th>Secondary level I</th>
<th>Secondary level II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Lower Degree</td>
<td>Intermediate Degree</td>
</tr>
<tr>
<td>German</td>
<td>2004</td>
<td>2004</td>
<td>2003</td>
</tr>
<tr>
<td>First foreign language (English, French)</td>
<td>-</td>
<td>2004</td>
<td>2003</td>
</tr>
<tr>
<td>Biology, Chemistry, Physics</td>
<td>-</td>
<td>-</td>
<td>2004</td>
</tr>
</tbody>
</table>
From standards to feedback

Output Domain (e.g., skills, cognitive processes)

Performance Standards: Structure and levels of domain-specific competence

Content Standards (e.g., state curriculum framework)

Content Domain
From standards to feedback

Evidential Aspects of Validation
(Procedural, Internal, and External Aspects)

Output Domain (e.g., skills, cognitive processes)

Item Domain

Performance Standards: Structure and levels of domain-specific competence

Content Standards (e.g., state curriculum framework)

Test Specifications

Test (Tasks /Item Bank)

Standard Setting (Cut Scores)
## Test domains

<table>
<thead>
<tr>
<th>Subject</th>
<th>Domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>German</td>
<td>reading, listening, orthography, language and reflection on language, writing</td>
</tr>
<tr>
<td>English/French</td>
<td>reading, listening</td>
</tr>
<tr>
<td>Mathematics</td>
<td>five main ideas (e.g., measurement, data and odds)</td>
</tr>
<tr>
<td>Biology</td>
<td>four competence domains (e.g., using scientific knowledge, generating scientific knowledge)</td>
</tr>
</tbody>
</table>
## Proficiency levels: English listening comprehension, intermediate degree (excerpts)

Based on the Common European Framework of Reference for Languages (2001)

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>V Optimal Standard (B2.2, C1)</strong></td>
<td>Can understand enough to follow longer input on unfamiliar and abstract topics. Understands a broad spectrum of idiomatic phrases and colloquial expressions. Can follow longer monologues and conversations even if they are not clearly structured. [C1]</td>
</tr>
<tr>
<td><strong>IV Norm Standard Plus (B2.1)</strong></td>
<td>Can understand the main propositions of standard input on concrete and abstract topics, even if content and language are complex. Can follow longer input and complex argumentations if the topic is to some extent familiar and it is structured by explicit signals. [B2.1]</td>
</tr>
<tr>
<td><strong>III Norm Standard (B1.2)</strong></td>
<td>Can understand factual information with low complexity on common everyday and work-related topics, is able to understand main propositions and individual pieces of information if the input is clearly articulated and the accent is familiar. [B1.2]</td>
</tr>
<tr>
<td><strong>II Minimal Standard (A2.2, B1.1)</strong></td>
<td>Can understand main points of clearly articulated standard input on familiar matters regularly encountered in school, leisure, etc. [B1.1]</td>
</tr>
<tr>
<td><strong>I Below Minimal Standard (A1, A2.1)</strong></td>
<td>Can understand expressions and words related to matters of immediate relevance. [A2.1]</td>
</tr>
</tbody>
</table>
From standards to feedback

Evidential Aspects of Validation (Procedural, Internal, and External Aspects)

Consequential Aspects of Validation (Utility and Impact Aspects)

Output Domain (e.g., skills, cognitive processes)

Performance Standards: Structure and levels of domain-specific competence

Content Standards (e.g., state curriculum framework)

Item Domain

Test Specifications

Test (Tasks / Item Bank)

Standard Setting (Cut Scores)

Test Use and Interpretations (Intended & Actual)

Test Impact

Content Domain
## Assessments in Germany

<table>
<thead>
<tr>
<th>Standards-based tests?</th>
<th>International</th>
<th>National IQB Assessments</th>
<th>School-level IQB Assessments (VERA)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PISA, PIRLS, TIMSS</td>
<td>Across the 16 states</td>
<td>Within the 16 states</td>
</tr>
<tr>
<td>Standards-based tests?</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Who is tested?</td>
<td>Sample (approx. 4,000-5,000)</td>
<td>Sample (approx. 30-40,000)</td>
<td>Population</td>
</tr>
<tr>
<td>Frequency</td>
<td>3 years, 5 years</td>
<td>3 years, 5 years</td>
<td>Every year</td>
</tr>
<tr>
<td>Main Purpose</td>
<td>System monitoring</td>
<td>System monitoring</td>
<td>School &amp; teaching improvement</td>
</tr>
<tr>
<td>Who is accountable?</td>
<td>Federal Ministry of Education; 16 State Ministries of Education</td>
<td>16 State Ministries of Education and their school authorities</td>
<td>Principals, teachers</td>
</tr>
</tbody>
</table>
## Assessments in Germany

<table>
<thead>
<tr>
<th>Standards-based tests?</th>
<th>International PISA, PIRLS, TIMSS</th>
<th><strong>National IQB Assessments</strong> Across the 16 states</th>
<th>School-level IQB Assessments (VERA) Within the 16 states</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standards-based tests?</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Who is tested?</td>
<td>Sample (approx. 4,000-5,000)</td>
<td>Sample (approx. 30-40,000)</td>
<td>Population</td>
</tr>
<tr>
<td>Frequency</td>
<td>3 years, 5 years</td>
<td>3 years, 5 years</td>
<td>Every year</td>
</tr>
<tr>
<td>Main Purpose</td>
<td>System monitoring</td>
<td>System monitoring</td>
<td>School &amp; teaching improvement</td>
</tr>
<tr>
<td>Who is accountable?</td>
<td>Federal Ministry of Education; 16 State Ministries of Education</td>
<td>16 State Ministries of Education and their school authorities</td>
<td>Principals, teachers</td>
</tr>
</tbody>
</table>
National IQB Assessments, 2nd Cycle: “IQB Educational Trends”

2009
Sprachliche Kompetenzen im Ländervergleich

2011
Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik
Ergebnisse des IQB-Ländervergleichs 2011

2012
IQB-Ländervergleich 2012
Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I

2015
IQB-Bildungstrend 2015
Sprachliche Kompetenzen am Ende der 9. Jahrgangsstufe im zweiten Ländervergleich

2016
IQB-Bildungstrend 2016
Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im zweiten Ländervergleich

Grade 9: German, English, French

Grade 4: German, Mathematics

Grade 9: Mathematics, Biology, Chemistry, Physics

© IQB, 2017
Changes in the proportion of students (grade 4) reaching at least the norm standard („Regelstandard“) between 2011 und 2016: mathematics
# Assessments in Germany

<table>
<thead>
<tr>
<th></th>
<th>International</th>
<th>National IQB Assessments</th>
<th>School-level IQB Assessments (VERA)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Across the 16 states</td>
<td>Within the 16 states</td>
</tr>
<tr>
<td>Standards-based tests?</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Who is tested?</td>
<td>Sample (approx. 4,000-5,000)</td>
<td>Sample (approx. 30-40,000)</td>
<td>Population</td>
</tr>
<tr>
<td>Frequency</td>
<td>3 years, 5 years</td>
<td>3 years, 5 years</td>
<td>Every year</td>
</tr>
<tr>
<td>Main Purpose</td>
<td>System monitoring</td>
<td>System monitoring</td>
<td>School &amp; teaching improvement</td>
</tr>
<tr>
<td>Who is accountable?</td>
<td>Federal Ministry of Education; 16 State Ministries of Education</td>
<td>16 State Ministries of Education and their school authorities</td>
<td>Principals, teachers</td>
</tr>
</tbody>
</table>

Standards-based tests?:
- **No** for International assessments (PISA, PIRLS, TIMSS)
- **Yes** for National IQB Assessments
- **Yes** for School-level IQB Assessments (VERA)
Goals of the school-level IQB assessments

- Implementation of measures to improve instruction
- VERA test results at school and class level
- Joint agreement on strategies for improvement
- Testing takes place in grades 3 and 8, 1-2 years before students are expected to meet the standards („alert system“).
- Analysis of results within the school (and with school supervision)
- Discussion of potential explanations for test results
VERA-feedback for schools (example) – approaches differ between states

### Reading Comprehension

<table>
<thead>
<tr>
<th>Grade</th>
<th>Task Description</th>
<th>Class</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Sie können in sehr vielschichtigen, umfangreichen Texten verstreute Informationen auffinden, verknüpfen und in anderen Zusammenhängen verwenden. Sie sind in der Lage, Interpretationsvorschläge plausibel zu beurteilen und in einem argumentativen Text zentrale Thesen herauszufinden. Sie können die Funktion einzelner Elemente einer Argumentationskette (z.B. Beispiele) und im Text enthaltene Wertungen erkennen.</td>
<td>31%</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Sie können in längeren Texten auch versteckte Informationen finden und verknüpfen sowie sich begründet für eine von mehreren vorgegebenen Interpretationen entscheiden. Sie sind in der Lage unter Berücksichtigung des gesamten Textes, Personen Merkmale zuzuordnen, auch wenn diese z.T. nicht ausdrücklich im Text benannt werden. Sie können erkennen, warum ein Erzähler einen Sachverhalt auf eine bestimmte Weise darstellt.</td>
<td>44%</td>
<td>29%</td>
</tr>
<tr>
<td>3</td>
<td>Sie können Informationen, die über Textabschnitte verteilt sind, miteinander verknüpfen und einfache Schlussfolgerungen ziehen. Vereinzelt können auch auf der Basis mehrerer Informationen Beweggründe für das Handeln zentraler Personen benannt werden.</td>
<td>44%</td>
<td>27%</td>
</tr>
<tr>
<td>2</td>
<td>Sie können Informationen, die über Textabschnitte verteilt sind, miteinander verknüpfen und einfache Schlussfolgerungen ziehen. Vereinzelt können auch auf der Basis mehrerer Informationen Beweggründe für das Handeln zentraler Personen benannt werden.</td>
<td>11%</td>
<td>10%</td>
</tr>
<tr>
<td>1</td>
<td>Sie können in kurzen Texten hervorgehobene, einzelne Informationen auffinden und diese mit einfachem Alltagswissen miteinander verknüpfen. Sie erkennen in einfach aufgebauten Texten das Hauptthema und können Wichtiges von weniger Wichtigem unterscheiden.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Proficiency Level 1
Proficiency Level 2
Proficiency Level 3
Proficiency Level 4
Proficiency Level 5
Student Assessment in Germany: Future Initiatives

- Feasibility study for computer-based assessment (starting with VERA).
- Development of modules for each test domain from which states/schools/teachers can choose (mandatory core module + optional modules for different levels of proficiency).
- Initiatives to improve usage of test results as a tool for developing instructional quality (important focus).
- Continued and improved use of large-scale assessments for further research (e.g., effects of different schooling models on students with special education needs, longitudinal addition to PISA 2012/IQB-Bildungstrend 2012).
Thank you for your attention!
Examples for educational standards: English listening comprehension, intermediate degree

Students are able to...

• follow the main points made in longer conversations,
• understand announcements on concrete issues that are spoken at normal speed in standard language,
• understand presentations if they are clearly structured and their complexity is limited and if students are familiar with the topic,
• understand the main information presented on the radio and on television regarding topics that are of personal interest.
Differences in the proportion of students (grade 4) reaching at least the norm standards in **2016** between the 16 states and Germany overall: **mathematics**