

Student Assessment

Session 2

Why student assessment is important

February 6, 2013

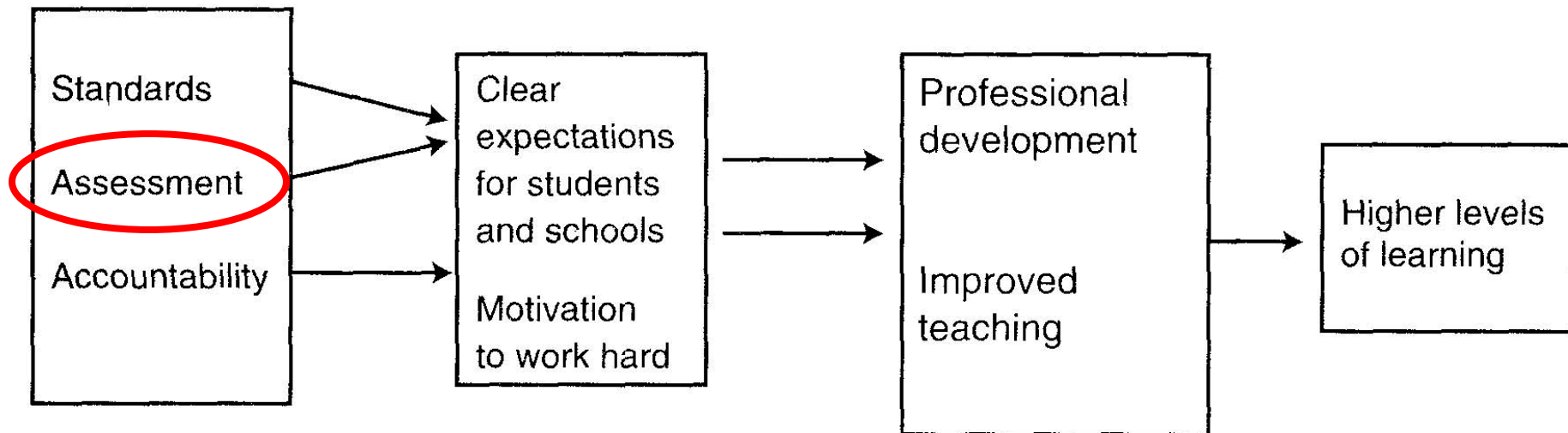
Room I2-220

Washington, DC





A theory of change



Session overview

- Key evidence and good practices for student assessment
- Different assessment models
- What matters most for student assessment systems?
 - Introduction to SABER-Student Assessment framework

Definitions, key evidence, good practices

What is student assessment?

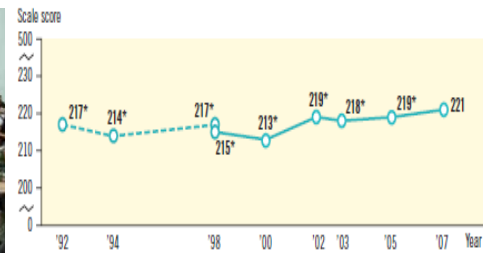
...the process of gathering and evaluating information on what students know, understand, and can do in order to make an informed decision about next steps in the educational process.

What is a student assessment system?

... a group of policies, structures, practices, and tools for generating and using information on student learning and achievement.



Figure 1. Trend in fourth-grade NAEP reading average scores



What counts as key evidence and best practices for effective assessment?



Classroom assessment



Strong link between high-quality assessment and better learning outcomes

Gains of half to full standard deviation on standardized tests (Black & Wiliam, 1998; Rodriguez, 2004)

More work needed to define and isolate characteristics that lead to improved outcomes (Bennett, 2011)

Examinations

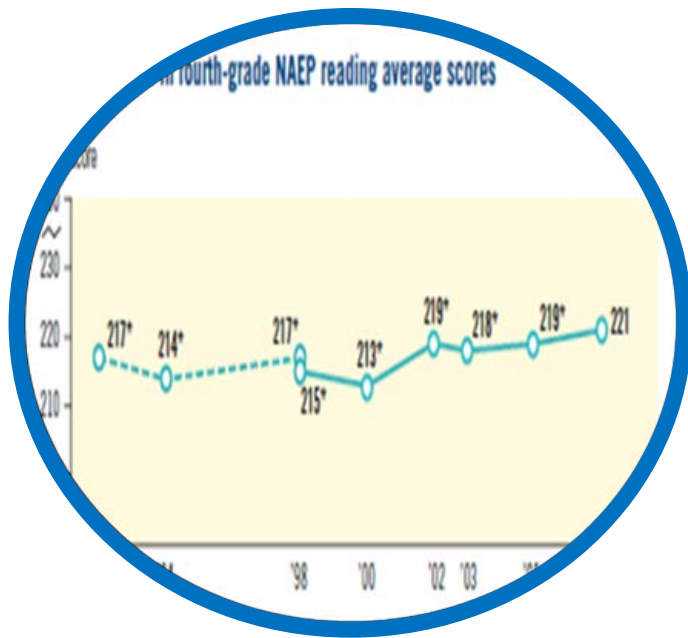


Link between exit exam policies and higher performance levels on international assessments (Bishop, Mane, & Bishop, 2001)

Link between specific characteristics of the test and student learning outcomes (Au, 2007; Hill, 2010)

Can have negative impact on students (Greaney & Kellaghan, 1995; Madaus & Clarke, 2001)

Large-scale 'survey' assessment



Simply reporting information about average school scores can lead to increased student performance (Hanushek & Raymond, 2003)

Overall weak, positive link between accountability uses of data and better learning outcomes (Carnoy & Loeb, 2002)

Much to learn about optimal design for accountability models that will produce best outcomes with fewest negative side effects

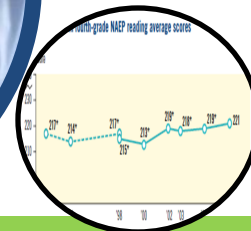
Different countries... different assessment models



Finland



- Strong emphasis on classroom-based formative assessment activities
- Limited use of high-stakes examinations
- Regular schedule of low-stakes, large-scale assessment activities for system monitoring





China

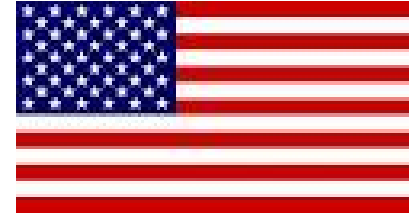


- Strong emphasis on high-stakes examinations
- Emergent focus on classroom assessment and large-scale assessment

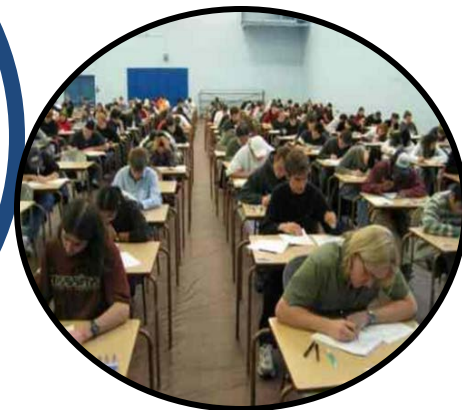
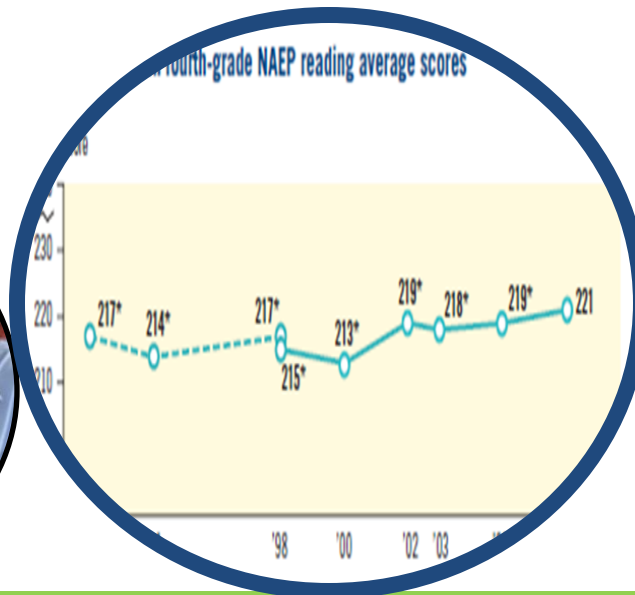




USA



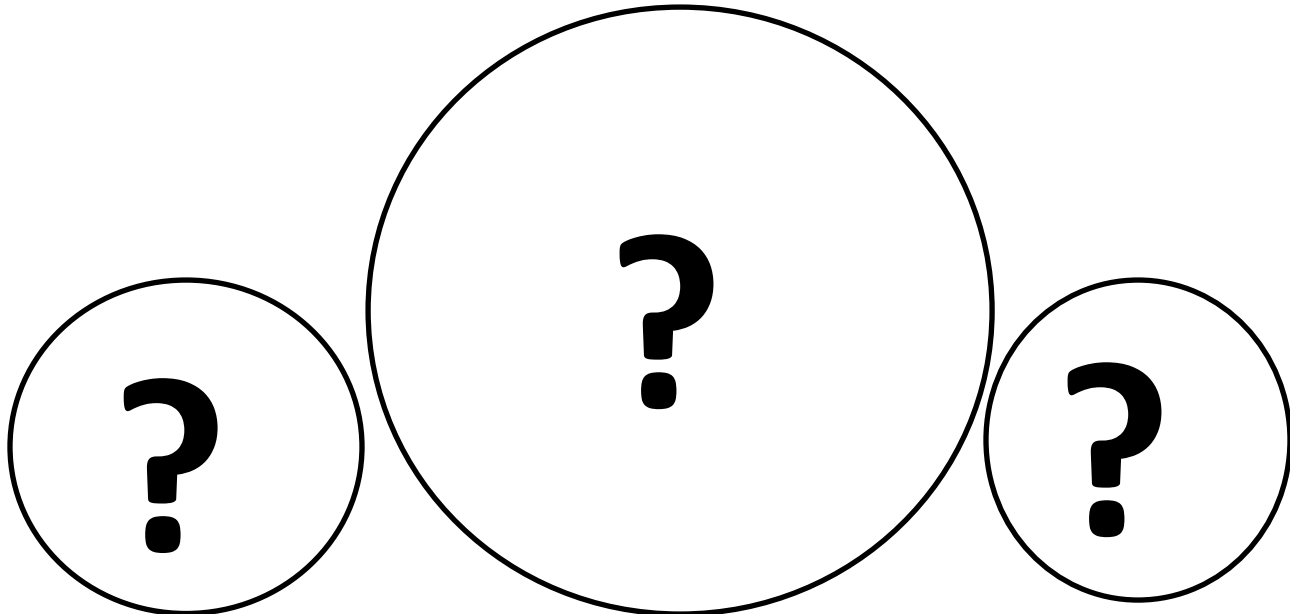
- Very strong emphasis on large-scale assessment
- Strong emphasis on classroom assessment and examinations





<Name of country in which you work>?

- Classroom assessment
- Examinations
- Large-scale assessment



Activity: Choosing among models

- Identify your preferred assessment model and discuss the reasons behind your preference.
 - Which is your preferred model?
 1. More emphasis on **classroom assessment**
 2. More emphasis on **external examinations**
 3. More emphasis on **large-scale assessments**
 - Why?
 - Under what conditions would you recommend it to a client?

What matters most for student assessment systems?

Assessment types/purposes

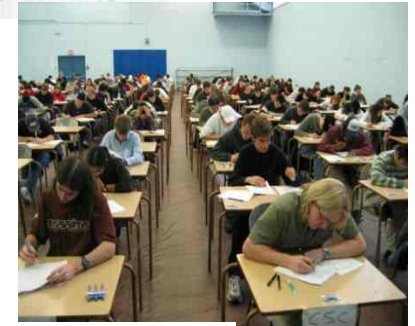
- **Classroom assessment**

- For improving teaching and learning



- **Examinations**

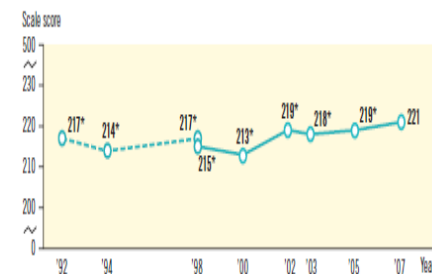
- For making high-stakes decisions about individual students



- **Large-scale assessment**

- For determining system learning levels and related factors

Figure 1. Trend in fourth-grade NAEP reading average scores



Quality drivers

- **Enabling context**

- Leadership, policies, institutional arrangements, human/fiscal resources



- **System alignment**

- Learning goals, curricula, opportunities

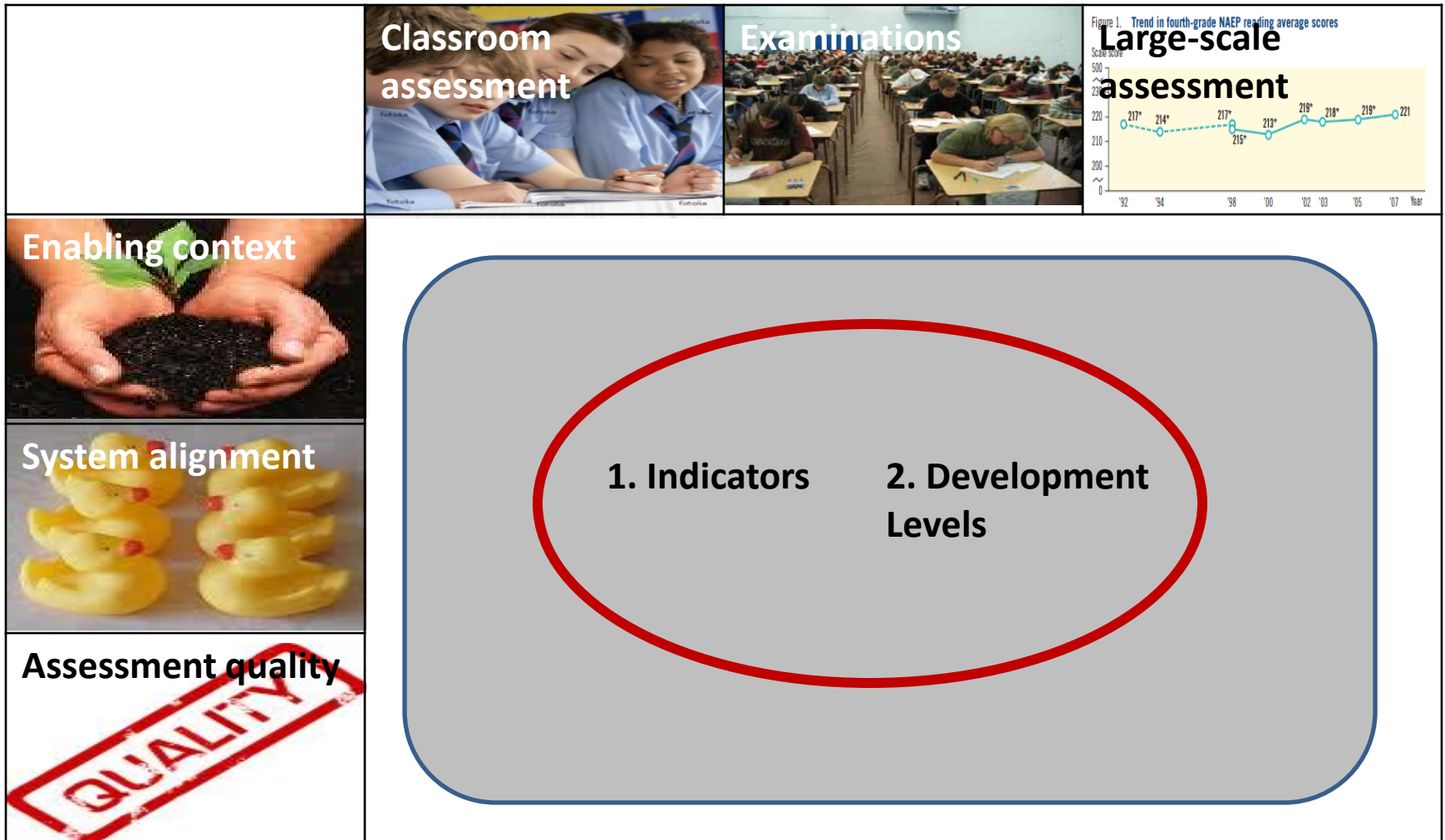


- **Assessment quality**

- Design, administration, analysis
- Reporting and use



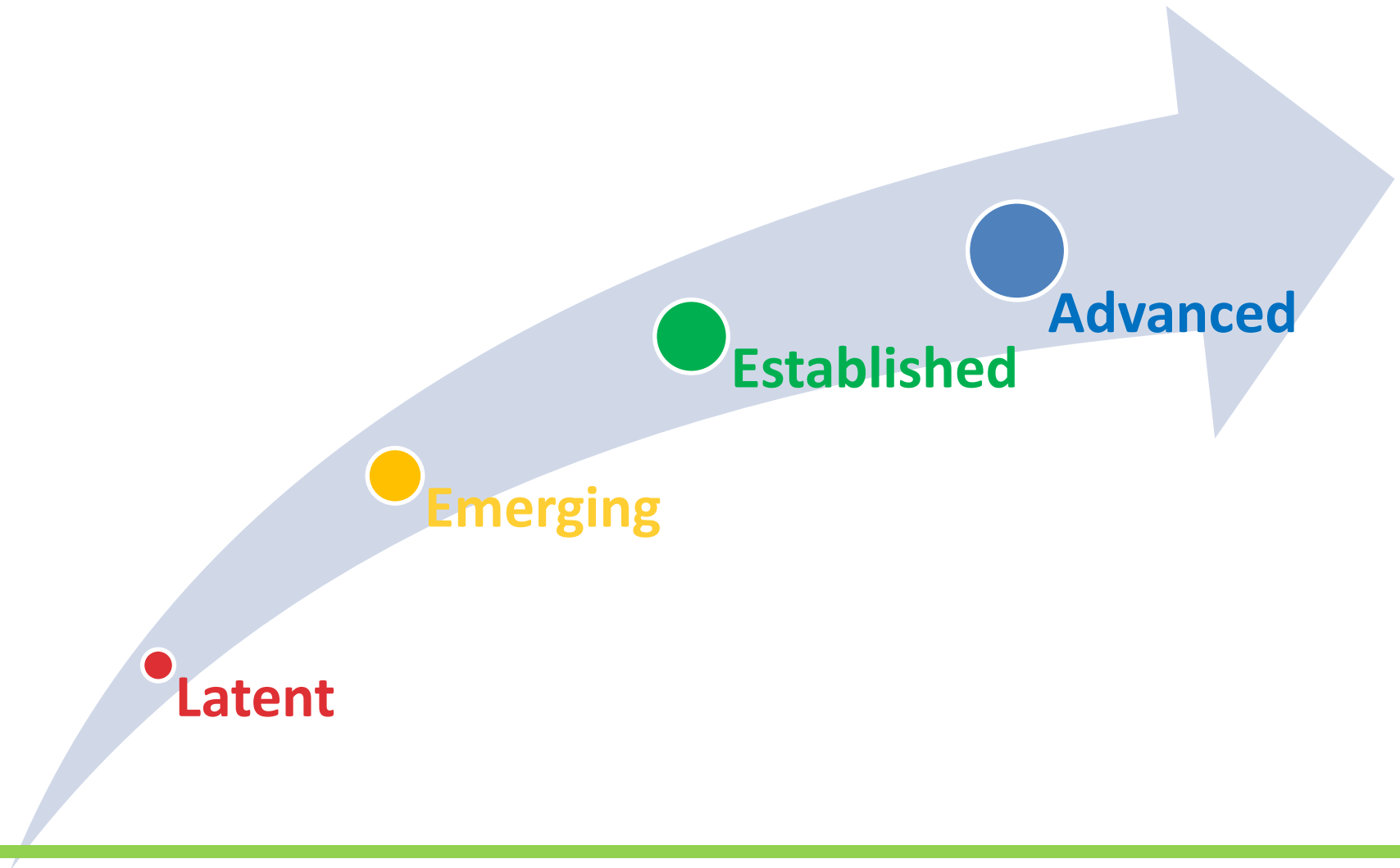
Framework



Indicators

- **Enabling context**
 - Policies, leadership, institutional arrangements, budget, human resources
- **System alignment**
 - Learning standards and goals, curricula
- **Assessment quality**
 - Technical features, dissemination, use, quality assurance

Development levels



Activity: Applying the framework to your regional context

- Using the framework as an organizing structure, discuss the relative strengths and needs of countries in the region where you work in the area of assessment.
- What data did you draw on to come to your conclusions?
- What other data would you need before taking any action?

Framework

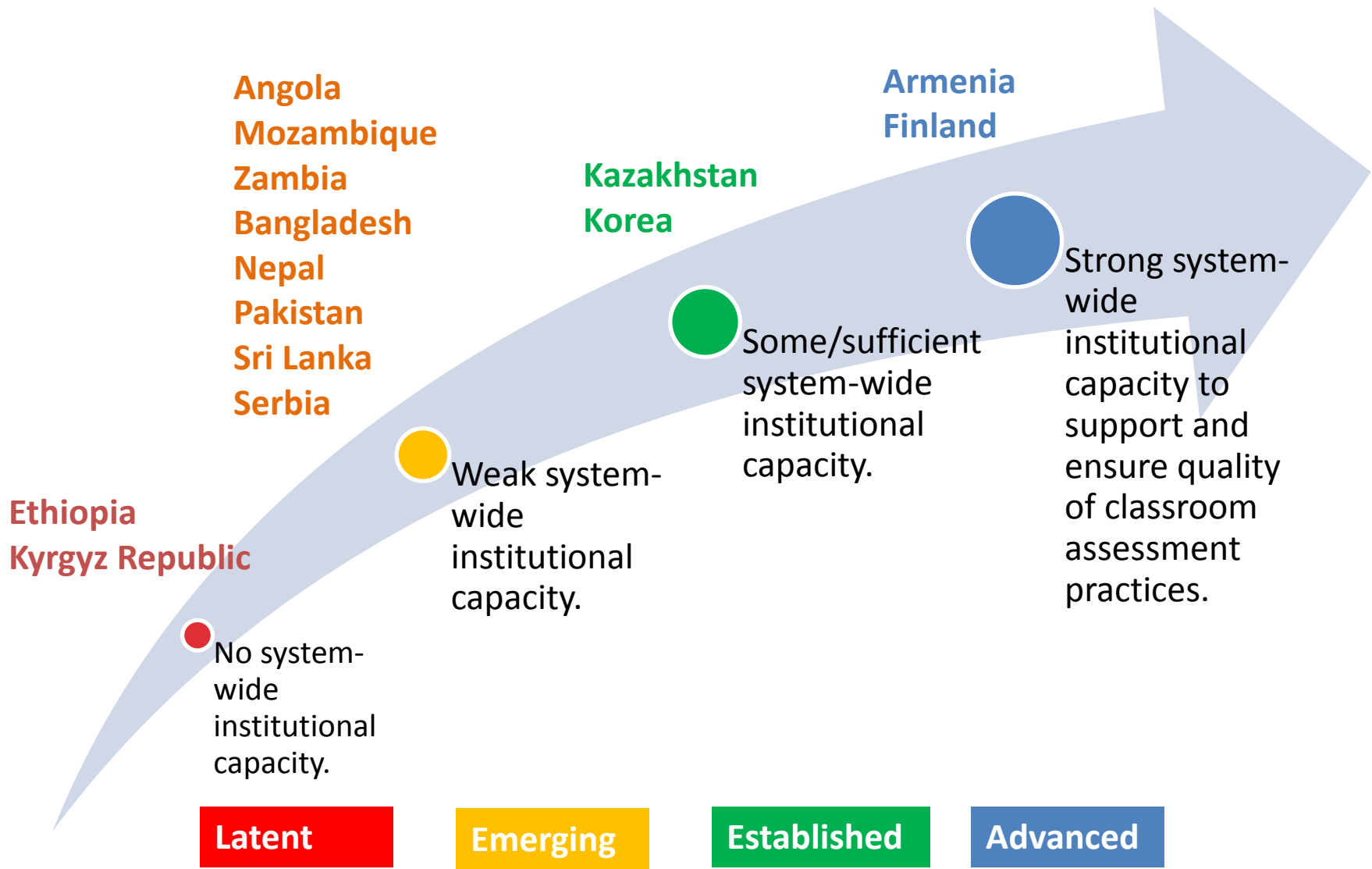
	 <p>Classroom assessment</p>	 <p>Examinations</p>	<p>Figure 1. Trend in fourth-grade NAEP reading average scores</p>  <p>Large-scale assessment</p> <table border="1"> <thead> <tr> <th>Year</th> <th>Score</th> </tr> </thead> <tbody> <tr><td>'92</td><td>217*</td></tr> <tr><td>'94</td><td>214*</td></tr> <tr><td>'98</td><td>217*</td></tr> <tr><td>'00</td><td>215*</td></tr> <tr><td>'02</td><td>219*</td></tr> <tr><td>'03</td><td>218*</td></tr> <tr><td>'05</td><td>219*</td></tr> <tr><td>'07</td><td>221</td></tr> </tbody> </table>	Year	Score	'92	217*	'94	214*	'98	217*	'00	215*	'02	219*	'03	218*	'05	219*	'07	221
Year	Score																				
'92	217*																				
'94	214*																				
'98	217*																				
'00	215*																				
'02	219*																				
'03	218*																				
'05	219*																				
'07	221																				
 <p>Enabling context</p>																					
 <p>System alignment</p>																					
 <p>Assessment quality</p>																					

Indicators

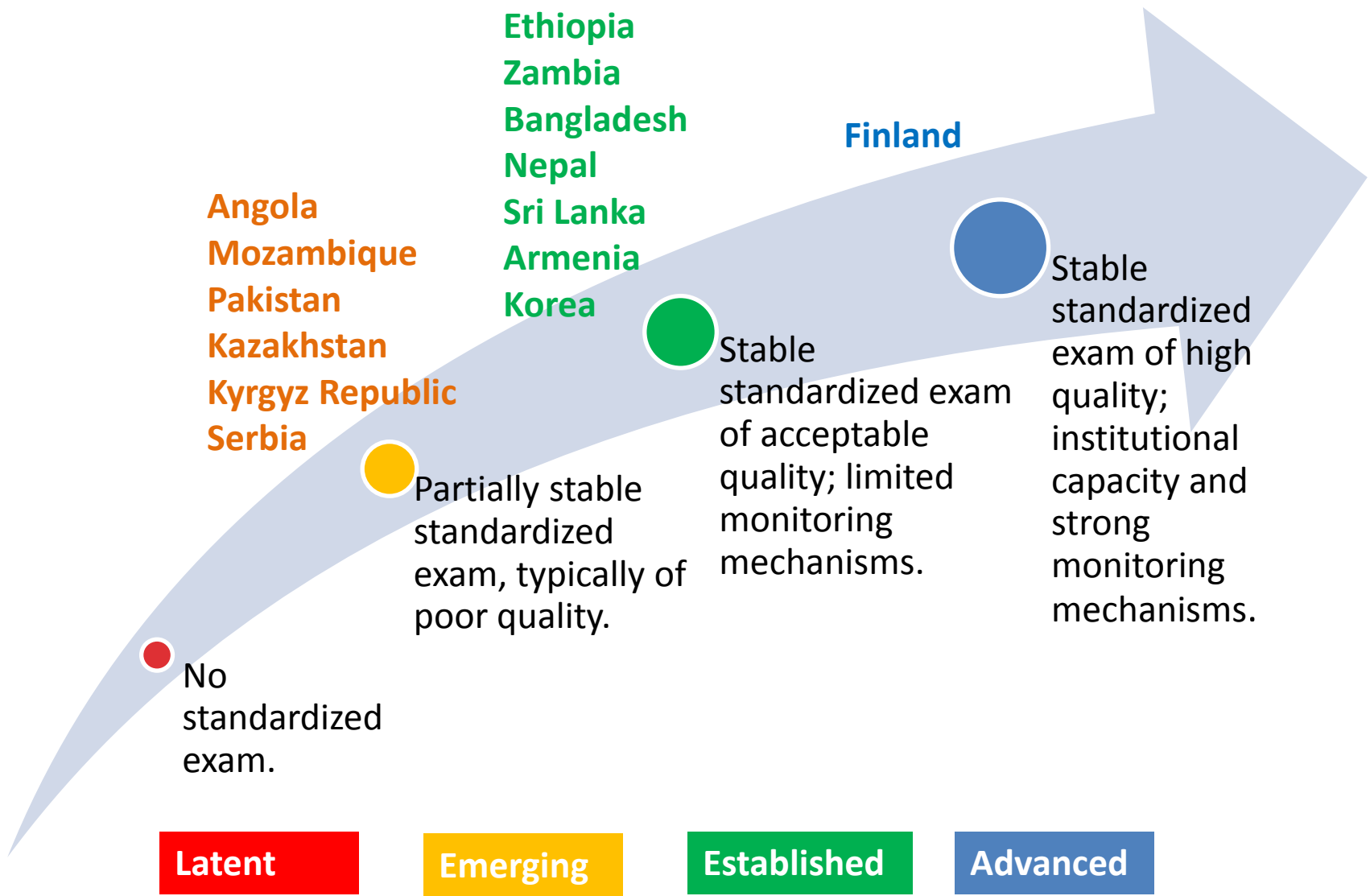
- **Enabling context**
 - Policies, leadership, institutional arrangements, budget, human resources
- **System alignment**
 - Learning standards and goals, curricula
- **Assessment quality**
 - Technical features, dissemination, use, quality assurance

Extra slides

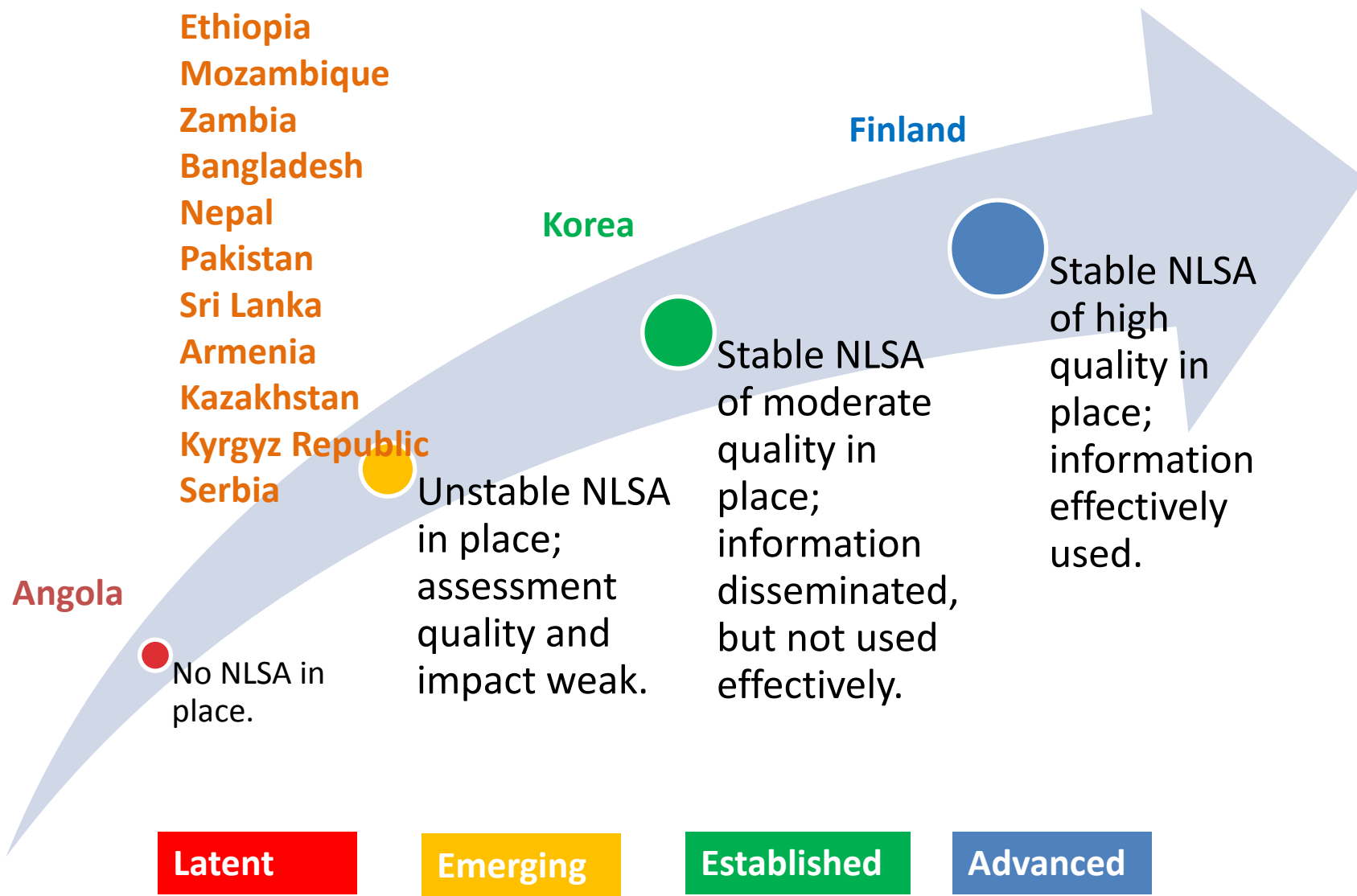
Classroom Assessment Development Levels



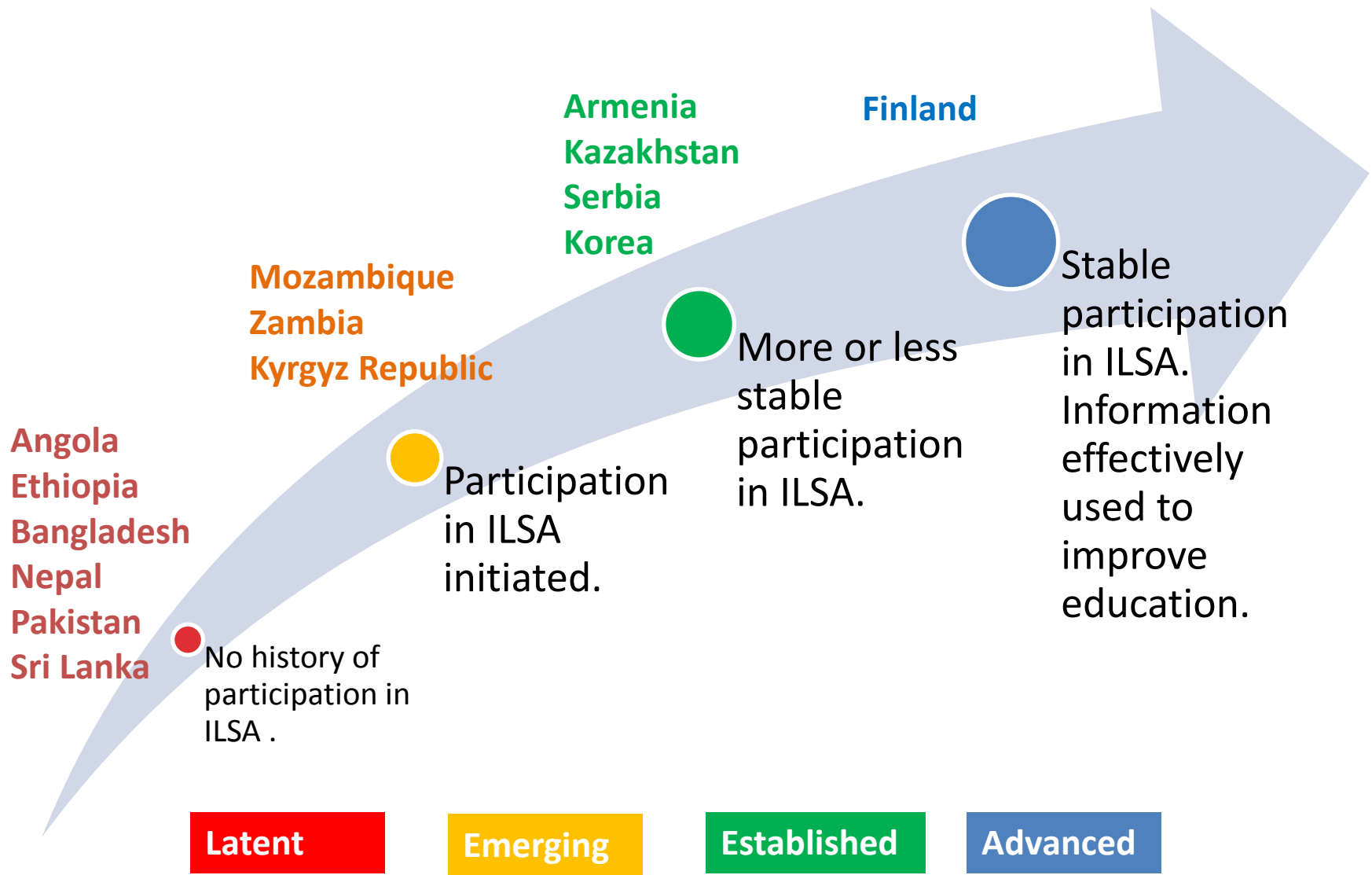
Examinations Development Levels



National Large-Scale Assessment Development Levels



International Large-Scale Assessment Development Levels



Sub-Saharan Africa

	Classroom Assessment	Examinations	Large-scale Assessment	
			National	International
Angola	Yellow	Yellow	Red	Red
Ethiopia	Red	Green	Yellow	Red
Mozambique	Yellow	Yellow	Yellow	Yellow
Zambia	Yellow	Green	Yellow	Yellow

South Asia

	Classroom Assessment	Examinations	Large-scale Assessment	
			National	International
Bangladesh	Yellow	Green	Yellow	Red
Nepal	Yellow	Green	Yellow	Red
Pakistan (Punjab & Sindh)	Yellow	Yellow	Yellow	Red
Sri Lanka	Yellow	Green	Yellow	Red

Eastern Europe, Central Asia

	Classroom Assessment	Examinations	Large-scale Assessment	
			National	International
Armenia	Blue	Green	Yellow	Green
Kazakhstan	Green	Yellow	Yellow	Green
Kyrgyz Republic	Red	Yellow	Yellow	Yellow
Serbia	Yellow	Yellow	Yellow	Green

Key trends and issues

- Blended cognitive and psychometric models
- Blurring of lines between classroom and external assessment events
- Measuring 21st century skills, cognitive, non-cognitive
- Key role of technology
- Key role of teachers
- Accountability uses of assessment data
 - Value-added models
 - NGOs and community mobilization

21st century skills

WAYS OF THINKING

Creativity and innovation
Critical thinking, problem solving, decision making
Learning to learn

TOOLS FOR WORKING

Information literacy
ICT literacy

WAYS OF WORKING

Communication
Collaboration

WAYS OF LIVING IN THE WORLD

Citizenship
Life and career
Personal and social responsibility

Sample item

Problem 1 How do different payload masses affect the altitude of a helium balloon?

Design Experiment **Run Experiment** **Interpret results**

Choose Values Make Prediction TRY IT! Make Table Make Graph Draw Conclusion

Altitude (feet)	Balloon Volume (cubic feet)	Time to Final Altitude (minutes)
36211	3083	36

Payload Mass (pounds)	Amount of Helium (cubic feet)
10	2275

Glossary Science Help Computer Help Next