

International Large-Scale Assessment Conference

Educational Testing Service
Princeton, NJ
March 16-18, 2011



Setting a Context

- Until relatively recently, educational data were not collected in a consistent or standardized manner.
- In 1958, a group of scholars representing various disciplines met at UNESCO in Hamburg, Germany to discuss issues surrounding the evaluation of schools and students through the systematic collection of data relating to knowledge, skills and attitudes.
- Their meeting led to the development of a feasibility study of 13 year olds in 12 countries covering 5 content areas and the formation of IEA in 1967.

Setting a Context

- Here in the United States in 1963, the Commissioner of Education, Francis Keppel, invited Ralph Tyler to develop a plan for the periodic assessment of student learning.
- Planning meetings were held in 1963 and 1964 and a technical advisory committee was formed in 1965.
- In April 1969, NAEP first assessed in-school 17 year olds in citizenship, science and writing.

Setting a Context

- Prior to IEA studies and NAEP, there were no assessment programs to report on students or adults as a group.
- The primary focus of assessments had been on measuring individual differences in achievement rather than on students' learning.
- And, the data that were collected dealt primarily with the inputs to educational systems rather than the yield of education.

Setting a Context

- Tyler's vision for NAEP was to focus on what groups of people know and can do rather than on what score an individual might receive on a test.
- NAEP assessments would be based on identified objectives with specifications determined by subject matter experts.
- Reports would be based on the performance of selected groups, not individuals, who responded correctly to the exercises.
- Assessment results were not based on grade-level norms.

Setting a Context

- Interpretations were limited to the set of items used in each assessment.
- In the 1980s, programs beginning with NAEP began to use item response theory (IRT) to allow for the creation of scales and the broadening of inferences going beyond individual items in the assessment.
- New methodology involving marginal estimation was developed to optimize the reporting of proficiency distributions based on complex designs such as BIB spiraling.

Growth and Expansion

- Number of assessments
- Number of participating countries
- Populations that are surveyed
- Domains/Constructs that are measured
 - Cognitive and background
- Methodology
- Modes of assessment
- Overall interest and use of the data

Growth and Expansion

- ... not being satisfied with assertions or self reports
- ... in response to policy makers and researchers wanting to know more
- ... asking more challenging questions
- ... and creating both the need and opportunity for new methodological and technological developments

Growing Interest

Policy makers

Corporate leaders

Educational researchers

Practitioners

Platform developers

Economists

Thought leaders

Methodologists

Psychologists

Sociologists

In Summary

- Growing interest in large-scale assessments and the data they provide is reflected in the topics of this conference and its speakers.
- We hope you find the conversations both interesting and stimulating.