

DEVELOPING A NAEP INDICATORS FRAMEWORK: LESSONS FROM MAJOR INTERNATIONAL AND DOMESTIC EDUCATION INDICATOR REPORTS

EXECUTIVE SUMMARY

By Alan Ginsburg and Marshall S. Smith

Introduction

This is the first of two reports exploring the use of the background data collected by the Nation Assessment of Educational Progress (NAEP) to develop *key education indicators* at national, state, and urban district system levels. *Key indicators are statistics that regularly measure an important condition of education.* For example, NAEP can tie to its student achievement results the reporting of background conditions on: students' attitudes, motivation and excessive absenteeism; measures of teacher quality; and indicators of the nature of reading and math instruction (e.g., amount of instructional time).

The GAO identifies three broad purposes of indicators:

- Increase transparency and public awareness.
- Foster civic engagement and collaboration.
- Monitor progress, establish accountability for results, and aid decision-making.

In a NAEP context, indicators also serve to:

- Identify for each cognitive assessment (e.g. reading) a set of key indicators, which are derived from the background variables and are monitored regularly.

Specifically, this first report is intended to develop a general indicators framework specifying an organizing structure, potential indicators, measurement criteria and reporting design. This report is based on a review of several major international or domestic data collections and reports produced by organizations other than NAEP:

International

- OECD's *Education At a Glance*
- International Education Association's 2011 TIMSS *Mathematics Assessment* covering grade 4 and 8.
- OECD's 2009 *PISA Report*

Domestic

- National Center for Education Statistics *Condition of Education*
- Education Week's *Quality Counts*
- U.S. Department of Education's *Annual Priority Performance Goals*
- National Academy of Sciences' *Key National Education Indicators*

Potential Indicators by Organizing Structure

Exhibit EX-1

Potential NAEP Education Indicators From Which To Select Key Indicators For K-12			
Locus of Education Activity	Key Drivers		
	- 21 st Century Skills - Common Core Standards - Instructional Technology		
	Results	Enablers	Context/Constraints
Student	<ul style="list-style-type: none"> • Command of core content, using NAEP scores • College readiness levels by age and grade • Career readiness (21st century skills) 	<ul style="list-style-type: none"> • Attended preprimary education • Chronic absenteeism • Student motivation and belief that hard work is more important than luck • Student positive attitudes toward subject • Student uses research-based approaches to learning subject • Student respect for teacher and visa versa • Participation in extra-curricular activities including community service 	<ul style="list-style-type: none"> • Home learning environment • Formal and informal learning outside school – nature of the their neighborhood
Teacher	<ul style="list-style-type: none"> • Proportion of teacher evaluations that distinguish them from a basic standard • Quality of work that the students have • Teachers spend time supporting other teachers 	<ul style="list-style-type: none"> • Teachers with less than 3-years experience • Teachers with mastery-level and current knowledge of content they are teaching. • Teachers with mastery-level and cotemporary knowledge of child and adolescent development • Teacher-student interactions that demonstrate high levels and qualities of involvement, language, stimulation, and expansion of thinking and cognition, and sensitivity to students' perspectives, individual experiences, and backgrounds • Teacher student interactions that indicate that teachers respect students. 	<ul style="list-style-type: none"> • Teacher working conditions • Average district teacher salary • Time teachers spend teaching • Teacher has high quality professional development and comprehensive induction programs Quality of the principal • Teachers belong to professional learning communities
School/ Classroom	<ul style="list-style-type: none"> • School subject area assessment outcomes • School performance rating/ranking within their state • Parent satisfaction (on surveys) • Completion rates from each kind of school – elementary to middle, middle to high, high to graduate, graduate to college or job? 	<ul style="list-style-type: none"> • Content of instruction aligned with standards • Effective use of technology to support instruction • School Climate – whether the school is a learning organization – do teachers work together? • Instructional time per subject • Engaged instruction in subject • Emphasis on continuous improvement on outcomes through both formative and summative assessments aligned with standards • Emphasis on continuous improvement of practices of teaching 	<ul style="list-style-type: none"> • School SES Composition • Safe & orderly school climate • Teacher-student ratio • School resource shortages • School lacks key characteristics, coaches for teachers, support systems for students, technology, books
System (district, state or nation)	<ul style="list-style-type: none"> - System core content outcomes 	<ul style="list-style-type: none"> • Support for implementation of new content standards • Alignment of assessment with content standards • Accountability with emphasis on continuous improvement 	<ul style="list-style-type: none"> • K-12 education spending as a share of gross domestic product • K-12 spending per student • Disparity in resources across districts within states

The indicator structure in Exhibit EX-1 is focused primarily around variables at the student, teacher, school/classroom and system levels that support learning outcomes

across three aspects of education conditions:

- *Results* indicators include student assessment outcomes (such as from NAEP), but also teacher evaluations that reflect student outcomes, and other outcomes such as secondary school completion and parent satisfaction with the school.
- *Enabler indicators* reflect formal learning at different levels of education. These include student exposure to preschool, teachers' knowledge and skills and their ability to apply them to create a challenging and supportive classroom learning environment; and school instructional time and student engagement in the content areas. Enablers also include system policies and regulations at district, state and national levels regarding teacher certification, standards, assessment, and accountability.
- *Context/constraint indicators* reflect factors not readily manipulable by the education system, although conditions may be changeable with proper interventions, such as schools intervening in the home learning environment. These factors include: learning at home and outside the school in formal and informal settings; factors influencing teacher quality including salaries and working conditions; and factors affecting the school learning environment including school safety, climate and class size.

Indicator Measurement

A sound measure for an indicator should meet criteria of validity, reliability, and consistency overtime.

Validity of Indicators. A valid measure is one that adequately captures the underlying education condition of interest. Combining responses from a number of questions around a topic into a larger comprehensive indicator scale can produce richer indicator measures than reporting on a single question, but this approach currently is not used in NAEP background factor analyses. Exhibit EX-2 illustrates a scale developed for TIMSS at grade 4 measuring students' early numeracy activities before beginning primary school.

Exhibit EX-2 Development of Indicator Scales from Multiple Questions

Exhibit 4.9: Early Numeracy Activities Before Beginning Primary School*

TIMSS 2011
Mathematics 4th Grade

Reported by Parents

Students were scored according to their parents' frequency of doing the six activities on the *Early Numeracy Activities* scale. Students **Often** engaged in early numeracy activities had a score on the scale of at least 10.3, which corresponds to their parents "often" doing three of the six activities with them and "sometimes" doing the other three, on average. Students **Never or Almost Never** engaged in such activities had a score no higher than 6.9, which corresponds to parents "never or almost never" doing three of the six activities with them and "sometimes" doing the other three, on average. All other students had parents who **Sometimes** engaged them in early numeracy activities.

Country	Often		Sometimes		Never or Almost Never		Average Scale Score
	Percent of Students	Average Achievement	Percent of Students	Average Achievement	Percent of Students	Average Achievement	
Hungary	75 (0.9)	528 (2.9)	23 (0.8)	495 (4.9)	1 (0.4)	--	11.1 (0.04)
Czech Republic	75 (0.8)	514 (2.3)	25 (0.8)	508 (3.6)	0 (0.1)	--	11.0 (0.03)
Slovak Republic	73 (1.0)	514 (3.3)	25 (0.8)	499 (5.2)	2 (0.5)	--	11.1 (0.05)
Northern Ireland	70 (1.2)	583 (3.5)	29 (1.2)	566 (4.9)	1 (0.2)	--	11.2 (0.05)
Russian Federation	69 (1.1)	547 (3.7)	29 (1.0)	533 (4.7)	2 (0.3)	--	10.9 (0.04)
Poland	68 (0.9)	488 (2.3)	31 (0.9)	471 (3.0)	1 (0.2)	--	10.8 (0.03)
Ireland	66 (0.9)	539 (2.9)	33 (0.8)	517 (3.7)	2 (0.3)	--	10.9 (0.04)
Australia	61 (1.2)	540 (3.7)	36 (1.2)	520 (4.1)	3 (0.4)	488 (13.4)	10.7 (0.05)
Austria	61 (0.9)	515 (2.6)	38 (0.9)	502 (3.5)	2 (0.2)	--	10.4 (0.03)
Croatia	60 (0.8)	496 (2.3)	39 (0.8)	482 (2.6)	1 (0.2)	--	10.5 (0.03)
Germany	59 (1.0)	538 (2.3)	40 (1.0)	528 (2.8)	2 (0.2)	--	10.4 (0.03)

Source: IEA, *TIMSS*, 2011

Reliability of Indicators. A reliable indicator measure is one that produces consistent results when repeatedly measuring the same underlying condition. Qualitative responses may be unreliable when sensitive to the position of the respondent. In the recently completed NAEP background paper on science, Exhibit EX-3 was presented showing that teachers were more likely than school principals to indicate that resources were "not at all available" within a school. This is not surprising as it is principals who are responsible for school resource availability.

Exhibit EX-3 Differences between teacher and school reported responses about science resource availability raise issues of response reliability

	Not at all	Small Extent	Moderate extent	Large extent
Science Kits are provided (teacher reported)	26	30	29	16
Science Kits are provided (school reported)	7	24	32	37
Science magazines and books are provided (teacher reported)	22	35	33	11
Science magazines and books are provided (school reported)	2	19	35	44

NAEP Data Explorer

Consistency over time. A consistent measure requires using the same measure for an indicator over time. To the extent that measures are changed from time period to time period then it is unclear whether a change comes about because of a real change in the underlying condition or changes in the measure. The report by the Expert Panel on Strengthening NAEP Background Questions (2012) addressed this issue in its recommendation 1d:

“Use consistency over time as a criterion to consider for question selection and wording. NAEP’s inconsistent inclusion of background questions weakens its potential to track trends and improvements within a subject area and topic.

For example, the Expert Panel found that only one-third of the 2011 questions asking about course offerings provided at least a 6-year trend. None of the questions about curriculum or school resources used in 2011 were found on the 2005 or earlier questionnaires.

Sources of Indicator Data. The reports that were studied use two ways to obtain indicator data, which differentiate them from NAEP.

First, TIMSS and PISA both conduct a household survey to obtain information directly from parents or guardians about socio-economic status and the home learning environment. TIMSS innovatively combined with PIRLS to develop a joint sample household survey for grade 4 students. The household survey included questions about:

- Early numeracy activities in the home before beginning primary school (See Exhibit EX-2)
- Early literacy activities in the home before beginning primary school
- Amount of exposure to preschool
- Family perception about child’s literacy and numeracy skills before entering primary school
- Family interaction with the child about school work
- Family perceptions about school
- Family literacy environment
- Family SES

A second source of data that is different from NAEP is the pooling of information across different surveys. Both the *Condition of Education* and *Education at a Glance* are drawn almost entirely from data series generated by other surveys. *Quality Counts* is a state-level amalgam of direct analyses of state policies by Education Week combined with data from other surveys, which prominently features NAEP assessment results.

A form of pooling could be the aligning of NAEP survey questions with international assessment items, as illustrated in Exhibit EX-4. The exhibit suggests that at least for U.S. middle schools, only about 12 percent of U.S. principals are having at least some difficulty filling vacancies for mathematics teachers. This compares with other Western

English-speaking countries of 41 percent of the principals having difficulty hiring math teachers in Australia, 37 percent in England, and 44 percent in New Zealand. Adding the same question about vacancies to the NAEP principal survey for mathematics

Exhibit EX-4
Schools Having Difficulties Filling Vacancies With Mathematics Teachers, Grade 8

Country	No Vacancies	Vacancies Are Easy To Fill	Vacancies Are Somewhat Difficult To Fill	Vacancies Are Very Difficult to Fill	Total of Vacancies Somewhat or Very Difficult To Fill
	Percent of students	Percent of students	Percent of students	Percent of students	Percent of students
Australia	25	34	31	10	41
Chinese Taipei	46	44	10	1	11
England	28	35	27	10	37
Finland	42	46	10	1	11
Hong Kong SAR	48	44	8	0	8
Japan	82	6	8	3	11
Korea, Rep. of	67	16	15	2	17
New Zealand	30	27	38	6	44
Norway	38	40	20	2	22
Russian Federation	81	11	6	2	8

would yield U.S. state-by-state comparisons.

Next Steps: Using the International and Domestic Indicator Framework as a Guide, Develop a NAEP Education Indicators Framework and Provide Examples with Current Data

A second report will be prepared for NAGB by December 2013 with a recommended set of *Key Indicators* and recommended improvements in NAEP data to strengthen indicator measurement or fill indicator gaps. The December 2013 report will:

- Specify a NAEP Education Indicators Framework for the background variables applicable across cognitive assessments.
- Propose indicators that are research-based and estimable by:
 - offering examples using current NAEP data.
 - suggesting changes to the current NAEP questionnaires.
 - introducing a fundamentally new NAEP questionnaire or drawing data from education surveys other than NAEP.
- Explore opportunities for combining NAEP with International or other NCES indicator-supporting data.
- Explore how NAEP reports could best display a pyramid information approach along the lines of an indicator dashboard.

DEVELOPING A NAEP INDICATORS FRAMEWORK: LESSONS FROM MAJOR INTERNATIONAL AND DOMESTIC EDUCATION INDICATOR REPORTS

An Exploratory Analysis Prepared for the National Assessment Governing Board

By Alan Ginsburg and Marshall S. Smith

August 2013

*Alan Ginsburg is an education consultant and analyst. He is former director of policy and program evaluation services for the U.S. Department of Education. **Marshall S Smith** is visiting Scholar, Carnegie Foundation for the Advancement of Teaching and former U.S. Under Secretary of Education. **The data analyses and interpretations in this report are those of the authors and do not necessarily represent the views of the National Assessment Governing Board.***

Table of Contents

<u>DEVELOPING A NAEP INDICATORS FRAMEWORK: LESSONS FROM MAJOR INTERNATIONAL AND DOMESTIC EDUCATION INDICATOR REPORTS</u>	3
<u>1. INTRODUCTION</u>	3
<u>2. INDICATOR BACKGROUND AND METHODOLOGICAL CONSIDERATIONS</u>	4
<u>3. INTERNATIONAL AND NATIONAL INDICATOR REPORTS</u>	7
INTERNATIONAL: OECD'S EDUCATION AT A GLANCE	7
INTERNATIONAL: IEA'S TIMSS (TRENDS IN INTERNATIONAL MATHEMATICS AND SCIENCE STUDY)	11
INTERNATIONAL: OECD'S PISA (PROGRAM FOR INTERNATIONAL STUDENT ASSESSMENT)	14
DOMESTIC: U.S. DEPARTMENT OF EDUCATION'S PRIORITY PERFORMANCE GOALS	18
DOMESTIC: EDUCATION WEEK'S QUALITY COUNTS	19
DOMESTIC: NCES'S CONDITION OF EDUCATION	21
DOMESTIC: NATIONAL ACADEMY OF SCIENCES' KEY NATIONAL EDUCATION INDICATORS	22
IMPLICATIONS	26
<u>4. PROPOSED DESIGN OF KEY INDICATORS FRAMEWORK</u>	27
INDICATOR ORGANIZING STRUCTURE	27
INDICATOR SELECTION	29
INDICATOR MEASUREMENT	32
SOURCES OF INDICATOR DATA	36
REPORTING ON INDICATORS	38
<u>5. NEXT STEPS: USING THE INTERNATIONAL AND DOMESTIC INDICATOR FRAMEWORK TO GUIDE DEVELOPMENT OF A NAEP INDICATOR FRAMEWORK AND PROVIDE EXAMPLES WITH CURRENT DATA</u>	39
<u>REFERENCES</u>	41

The authors wish to thank Lawrence Feinberg, Assistant Director for Reporting and Analysis of the National Assessment Governing Board, for his many analytic and editorial contributions to strengthen this report.

DEVELOPING A NAEP INDICATORS FRAMEWORK: LESSONS FROM MAJOR INTERNATIONAL AND DOMESTIC EDUCATION INDICATOR REPORTS

1. Introduction

This is the first of two reports exploring the use of the background data collected by the National Assessment of Educational Progress (NAEP) to develop *key education indicators* at national, state, and urban district system levels. The report has been commissioned by the National Assessment Governing Board (NAGB) to follow-up on recommendations by its Expert Panel on Strengthening the NAEP Background Questions to “develop key background indicators.” (Smith, et.al., 2012).

Exploring NAEP’s potential as part of a national and sub-national system of key education indicators is a priority for enhancing the value of NAEP. Indicators are statistics that regularly measure an important condition of education (NRC, 2012). The regular (every two years) collection of background variables for NAEP cognitive assessments, has the unique potential to strengthen our understanding of the assessment results and expand the richness of statistics on the quality and equality of our educational systems at the national, state, and urban district levels. For example, NAEP can tie to its student achievement results the reporting of background conditions on students’ attitudes, motivation and excessive absenteeism, measures of teacher quality, and the nature of reading and math instruction, including the amount of instructional time at multiple levels of the education system.

Specifically, this first report explores a number of the major international and domestic data collections and reports produced by organizations other than NAEP. These data collection efforts, including those associated with international assessments of school achievement, specifically generate data organized around well-defined indicators or they produce and report data describing key components of the education system much like an indicators report.

The aim of this report is to distill from these other education data and reports an *indicator framework* that might help structure and guide similar work on NAEP. A NAEP indicators framework would specify key features of a useful and valid indicator system including its design. *Within the NAEP context, an indicators framework could identify for each cognitive assessment (e.g. reading) a set of key indicators, which are derived from the background variables and regularly monitored.*

A second report, due in December 2013, will apply this report's indicator framework and use current NAEP data to develop some of the indicators proposed. It will also identify priority areas where NAEP could collect additional background data to address indicator gaps, or draw from other NCES statistics collections to fill these gaps.

The following sources of indicator or large-scale data reports similar to an indicator report are explored in this first report:

International

- OECD's *Education At a Glance*
- International Education Association's *TIMSS Math Assessment* covering grade 4 and 8.
- OECD's 2009 *PISA Report*

Domestic

- National Center for Education Statistics *Condition of Education*
- Education Week's *Quality Counts*
- U.S. Department of Education's *Annual Priority Performance Goals*
- National Academy of Sciences' *Key National Education Indicators*

The examination of these sources of indicator frameworks proceeds as follows:

- Section 2: Indicator Background and Methodological Considerations
- Section 3: International and National Indicator Reports
- Section 4: Proposed Design of Key Indicators Framework
- Section 5: Next Steps: Using the International and Domestic Indicator Framework to Guide Development of a NAEP Education Indicator Framework

2. Indicator Background and Methodological Considerations

The design of an education indicators framework begins with identifying the different purposes of education indicator systems. The focus must be on key national education indicators that are the ones most important to track overtime. A key indicator system is thus distinguished from a statistical effort, such as the NCES *Digest of Education Statistics* (<http://nces.ed.gov/Programs/digest/>), which attempts to produce an inclusive volume describing many aspects of the education system in some detail.

Within the scope of a key national indicator system, the Government Accountability Office (GAO) has identified three main purposes:

- *Increase transparency and public awareness* by giving decision-makers and the public easy, single-source access to credible and relevant information.
- *Foster civic engagement and collaboration* by bringing “increased pressure to bear on diverse parties in the public and the private sectors...to break down traditional boundaries between various actors and organizations” (GAO, 2011).
- *Monitor progress, establish accountability for results, and aid decision-making.* This focus includes assessing performance, discussing options, and making choices about major education concerns of fundamental and long-term significance.

This third purpose of monitoring progress and establishing accountability for results has been a particularly prominent aim of federal government-wide performance indicator policy including performance indicators established by the U.S. Department of Education. When an education indicator is associated with a defined policy or education objective, it becomes *an education performance indicator* that over time measures progress toward that objective. At the federal level, the Government Performance and Results Act (GPRA) requires every federal department, including the U.S. Department of Education, to establish performance indicators for every program. These indicators have been required since 1997.

Currently, as part of its GPRA reporting, the U.S. Department of Education has established for FY13 “priority performance goals” that illustrate the strategic importance of indicator measurement of educational conditions that might influence achievement. For pre-K through secondary school, the Education Department’s priority performance goals include:

- *“Improve outcomes for all children from birth through third grade.* By September 30 2013, at least nine states will implement a high-quality plan to collect and report disaggregated data on the status of children at kindergarten entry.
- *Improve learning by ensuring that more students have an effective teacher.* By September 30th, 2013, at least 500 school districts will have comprehensive teacher evaluation and support systems and a majority of States will have statewide requirements for comprehensive teacher and principal evaluation and support systems.
- *Demonstrate progress in turning around the nation's lowest-performing schools.* By September 30th 2013, 500 of the nation's persistently lowest-achieving schools will have demonstrated significant improvement and serve as potential models for future turnaround efforts.

- *Make informed decisions and improve instruction through the use of data.* By September 30th, 2013 all states will implement comprehensive statewide longitudinal data systems.
- *Prepare all students for college and career.* By September 30th 2013, all states will adopt internationally-benchmarked college- and career-ready standards.”

NAEP background variable data could potentially provide independently generated indicators in most of these priority areas that would add information value in several ways.

One important way NAEP would add value is through its capacity to *disaggregate national data* to provide comparable indicators across all states and more than 20 urban districts. In this context, NAEP started out reporting only nationally representative assessment data. Following the report, *A Nation At Risk* (1983), then Secretary of Education Ted Bell wanted a way to hold States accountable for improving measurable performance of student outcomes. At his direction, the Department of Education staff developed and published a Wall Chart, a one-page summary set of less-than-perfect state-by-state indicators of student outcomes, education services and context (student characteristics). Between 1984 and 1989 the announcement of the Wall Chart annually produced the Department’s largest press conferences.

But methodological limitations of the available State-by-State data led Congress, heavily supported by the Council of Chief State School Officers, to fund in 1988 a larger NAEP sample representative state-by-state (Ginsburg, Noell, and Plisko, 1988). NAEP is unique among NCES databases in yielding *comparable student assessment data for math, reading or science* subjects coupled with student, teacher and school background data for every state and currently for 21 individual urban districts.

A second way NAEP could potentially add value is *consistency of the measurement of the indicator series over time*. The real power of indicators comes from establishing baselines and then measuring change regularly over time. NAEP offers the potential for consistent measures from repeated administrations. Great care is taken to ensure comparability of NAEP assessments over time.

Unfortunately, as the Expert Panel report (2012) concluded, “NAEP’s inconsistent inclusion of background questions weakens its potential to track trends and improvements within a subject area and topic.” Nonetheless, more consistent and useful data series could be created from existing or new NAEP measures.

Along with different characteristics of individual indicators, *different perspectives of the education system can generate different frameworks* for identifying key indicators. One system perspective is by the *stages of student learning*. NAEP currently does not address preschool, but begins with a grade 4 assessment. An end-

of-preschool assessment has been discussed. Also, other surveys ask retrospective questions of parents or guardians as shown below.

A second system perspective is a *production function model*. This applies an economic model to education typically covering school contexts, school inputs, school services and student outcomes. Each of these components has multiple factors and research on the significance of each component for outcomes that can help distill the key indicators.

A third perspective is reflected in a National Academy of Sciences (1991) report *Education Counts*, which guides indicator development based on identifying six *critical issue areas* that an indicator system should address:

1. learner outcomes;
2. quality of educational institutions;
3. readiness for school;
4. societal support for learning;
5. education and economic productivity; and
6. equity (measures of resources, demographics, and students at risk).

Development of an indicators framework faces the challenge of defining a limited set of key education indicators or performance indicators within an overall system of indicators at different levels of disaggregation and over time. Specifying the features of the indicators framework is a critical initial step toward this prioritization. To move the indicator framework selection process forward, the next section explores key features related to indicators identified in the seven major international and national indicator or data reports listed above.

3. International and National Indicator Reports

This section examines seven international and domestic indicator-producing data systems and reports. Each system is described with respect to purpose, organizing framework and an overview of indicators focused around NAEP-relevant ages of early childhood and K-12 education.

International: OECD's Education at a Glance

This annual report draws on various OECD surveys to measure the current condition of education internationally. The indicators “provide information on the human and financial resources invested in education, how education and learning systems operate and evolve, and the returns to educational investments.”(OECD, 2013, p.17).

Exhibit 3-1. Education at A Glance Organizing Framework for Indicators			
Indicator Level of Actors of Education System	Indicator Focus		
	1. Education and learning outputs and outcomes	2. Policy levers and contexts shaping educational outcomes	3. Antecedents or constraints that contextualise policy
I. Individual participants in education and learning	1.I. The quality and distribution of individual educational outcomes	2.I. Individual attitudes, engagement, and behaviour to teaching and learning	3.I. Student learning conditions and teacher working conditions
II. Instructional settings	1.II. The quality of instructional delivery	2.II. Pedagogy, learning practices and classroom climate	3.II. Student learning conditions and teacher working conditions
III. Providers of educational services	1.III. The output of educational institutions and institutional performance	2.III. School environment and organisation	3. III. Characteristics of the service providers and their communities
IV. The education system as a whole	1.IV. The overall performance of the education system	2.IV. System-wide institutional settings, resource allocations, and policies	3. IV. The national educational, social, economic, and demographic contexts
Cross-cutting policy issues addressed: <ul style="list-style-type: none"> • Quality of educational outcomes and educational provision; • Equality of educational outcomes and equity in educational opportunities; and • Adequacy, effectiveness and efficiency of resource management. 			
Source: Adapted from <i>Education At a Glance (2013)</i> .			

Organizing framework. Exhibit 3-1 displays the three factors that form the basis for the organizing framework for *Education at a Glance 2013* (OECD, 2013).

- The left-hand column organizes indicators by the level of the “actors” in the education system. It assesses the functioning and impact of education systems at four levels contributing to overall learning outcomes. These include: I. Individual participants in education and learning; II. The instructional setting and the learning environment involving teachers and classrooms within institutions; III. The educational institutions (e.g., schools, informal education providers) that are providers of educational services; and IV. The national education system as a whole. In U.S. NAEP terms, the different levels approximately correspond to students, teachers, schools and classrooms, and the national, state and district systems.
- The columns cluster the indicators by whether the focus is on: 1. The education outcomes/outputs of the education system; 2. Policy levers and contexts which shape the outputs/outcomes; and 3. The antecedents that

define or constrain policy (e.g, teacher working conditions).

- At the bottom of the table, the cross-cutting policy issues indicate that each of the cells in the framework can be examined from three policy perspectives of quality, equality, and efficiency characteristics of resource management)

Indicators. The indicators reported in *Education at a Glance* are categorized in Exhibit 3-2 and are intended to emphasize one or more cells of this framework. The indicator structure resembles a production model with outputs, resources, access to education services and the learning process. With respect to a focus on early childhood or K-12 education, *Education at a Glance* covers:

- Chapter A, *Output of Educational Institutions and the Impact of Learning*, focuses only on students expected to complete upper secondary education.
- Chapter B, *Financial and Human Resources Invested in Education*, measures the amount of resources and types of expenditures on education at different levels with limited data on preprimary education.
- Chapter C, *Access to Education, Participation and Progression*, includes an important focus on access to early childhood education and transitions from secondary to tertiary education or from school to work.
- Chapter D, *The Learning Environment and Organization of Schools*, focuses on indicators of instructional time, teacher qualifications, teachers' working time and teacher salaries.

Looking across *Education at a Glance*, the Chapter A indicators relate largely to the first column. However, as *Education at a Glance* notes the indicators for the other chapters are a mixture of policy levers and contexts.

Exhibit 3-2. Education At a Glance Indicators

Chapter A. The Output of Educational Institutions and the Impact of Learning

Indicator A1 To what level have adults studied?

Indicator A2 How many students are expected to complete upper secondary education ?

Indicator A3 How many students are expected to complete tertiary education ?

Indicator A4 How many students complete tertiary education ?

Indicator A5 How does educational attainment affect participation in the labour market?

Indicator A6 What are the earnings premiums from education ?

Indicator A7 What are the incentives to invest in education ?

Indicator A8 What are the social outcomes of education ?

Chapter B. Financial and Human Resources Invested In Education

Indicator B1 How much is spent per student?

Indicator B2 What proportion of national wealth is spent on education ?

Indicator B3 How much public and private investment in education is there?

Indicator B4 What is the total public spending on education ?

Indicator B5 How much do tertiary students pay and what public support do they receive?

Indicator B6 On what resources and services is education funding spent?

Indicator B7 Which factors influence the level of expenditure on education ?

Chapter C. Access to Education, Participation and Progression

Indicator C1 Who participates in education ?

Indicator C2 How do early childhood education systems differ around the world?

Indicator C3 How many students are expected to enter tertiary education ?

Indicator C4 Who studies abroad and where?

Indicator C5 Transition from school to work: where are the 15-29 year-olds?

Chapter D. The Learning Environment and Organisation of Schools

Indicator D1 How much time do students spend in the classroom?

Indicator D2 What is the student-teacher ratio and how big are classes?

Indicator D3 How much are teachers paid ?

Indicator D4 How much time do teachers spend teaching?

Indicator D5 Who are the teachers?

Source: OECD, 2013

International: IEA’s TIMSS (Trends in International Mathematics and Science Study)

TIMSS 2011, like NAEP, administers mathematics and science assessments to grade 4 and 8 students. TIMSS 2011 gathered nationally representative samples of students in 63 countries and 14 benchmarking entities (including U.S. states). Fifty-two countries and seven benchmarking entities participated in the fourth grade assessment, and 45 countries and 14 benchmarking entities participated in the eighth grade assessment. While many European OECD countries did not participate in TIMSS when the more progressive PISA assessment was launched in 2000, the 2011 TIMSS survey involved many OECD countries including Finland, which is one of the highest scorers on PISA. Note that a companion survey to TIMSS, Progress in International Reading Literacy Study (PIRLS), assesses reading at grade 4. TIMSS and PIRLS collaborated on a common grade 4 household survey.

Exhibit 3- 3. TIMSS 2013 Questionnaires
<i>Student Questionnaires (grads 4 & 8):</i> basic demographic information, their home environment, school climate for learning, and self-perception and attitudes toward learning mathematics and science.
<i>Home Questionnaire</i> (grade 4 to homes of students participating in joint TIMSS & PIRLS survey) preparations for primary schooling, including attendance in preschool and literacy- and numeracy-centered activities in the home before the child began school, such as reading books, singing songs, or playing with number toys. Parents answered questions about home resources in addition to information about their highest level of education and employment situations.
<i>Teacher Questionnaires:</i> Asked students’ teachers about their education, professional development, and experience in teaching, coverage of the mathematics and science curriculum and about the instructional activities and materials used in the class.
<i>School Questionnaires:</i> Principals answered questions about student demographic characteristics, the availability of resources, types of programs, and environments for learning in their schools.
<i>Curriculum Questionnaires:</i> Answered by TIMSS 2011 National Research Coordinator questions centered on the organization and content of the curriculum in mathematics and science.
Source: TIMSS, 2011

TIMSS, like NAEP, is a general-purpose data collection from which indicators about key conditions of math and science education are derived. For purposes of indicator development, TIMSS includes contextual surveys that explore the “interplay of societal, school, and home environmental factors” on the achievement results.

The TIMSS contextual questionnaires (Exhibit 3-3) measure factors that affect student learning by administering questionnaires to students, parents or guardians (grade 4), teachers, schools, and the TIMSS national research coordinator (about the organization and content of the curriculum questionnaire). These questionnaires report on a variety major influences on student learning, including student engagement in classroom learning, home support for student learning, the scope and coverage of the mathematics or science curriculum, teacher preparation for mathematics or science instruction, and school resources and learning climate.

Exhibit 3-4. TIMSS & PIRLS Grade 4 Home Questionnaire Sample Items, 2011

Before your child began primary/elementary school, how often did you or someone else in your home do the following activities with him or her?

Check **one** circle for each line.

	Often	Sometimes	Never or almost never
a) Read books -----	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
b) Tell stories -----	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
c) Sing songs -----	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
d) Play with alphabet toys (e.g., blocks with letters of the alphabet) -----	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
e) Talk about things you had done -----	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
f) Talk about what you had read -----	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
g) Play word games -----	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
h) Write letters or words -----	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
i) Read aloud signs and labels -----	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
j) Say counting rhymes or sing counting songs -----	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
k) Play with number toys (e.g., blocks with numbers) -----	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
l) Count different things -----	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
m) Play games involving shapes (e.g., shape sorting toys, puzzles) -----	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
n) Play with building blocks or construction toys -----	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
o) Play board games or card games -----	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

How well could your child do the following when he/she began primary/elementary school?

Check **one** circle for each line.

	Very well	Moderately well	Not very well	Not at all
a) Recognize most of the letters of the alphabet -----	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
b) Read some words -----	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
c) Read sentences -----	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
d) Write letters of the alphabet -----	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
e) Write some words -----	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Source: TIMSS/PIRLS, Grade 4 Learning to Read Survey. Available July 2013 online: http://timssandpirls.bc.edu/timss2011/downloads/TP11_HQ.pdf.

An extensive *home questionnaire* for the parents of students at grade 4 is of particular note in the 2011 TIMSS; there is no such survey in NAEP. The home questionnaire is part of a joint administration with the 2012 Progress in International Reading Literacy Study (PIRLS) for grade 4. It covers family well-being and learning opportunities and expectations for both math and reading. This contrasts with NAEP's reliance on a brief and somewhat unreliable student questionnaire for grade 4 that covers home environment and family socio-economic status.

An example of the comprehensive home context information TIMSS gathers is the questions that ask parents about education activities prior to the child entering primary school (Exhibit 3- 4). The left-hand column asks a parent/guardian to

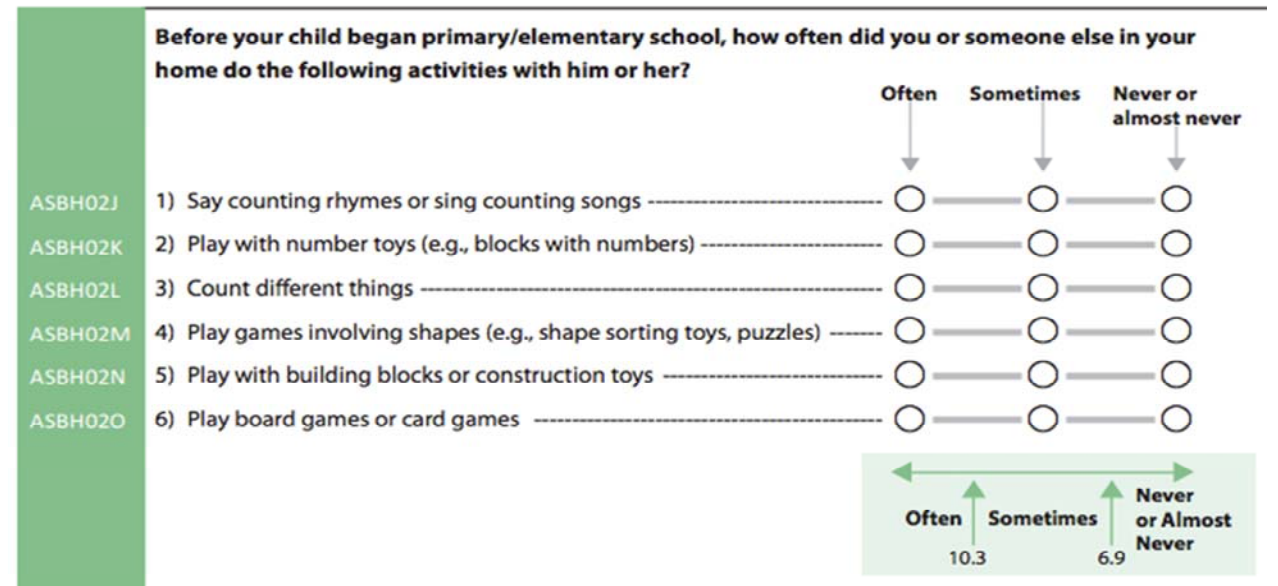
report on the frequency of various learning activities they did with their grade 4 child before the child began school. The right hand column asks about their child's reading ability before entering school. There are similar questions for mathematics. These data are subject to the normal error of retrospective responses, but the TIMSS analyses indicate they have predictive validity with current TIMSS scores.

Additional home questions inquire about current activities of the family in helping their child with math, reading and homework. There are also questions to the responder at the child's home about their perceptions of a child's school, personal reading habits, books in the home and how the child and parent talk together. Socio-economic status measures include questions about education, occupation, and income that are not accurately obtained from a fourth grader. TIMSS grade 8 responses about the home are from a student questionnaire.

A second feature of TIMSS, which differentiates it from NAEP's treatment of background variables, is essential for indicator development: TIMSS creates *numeric scales* by combining answers to questions focused around a similar background condition. Because a single question rarely captures the richness and multi-dimensionality characterizing an education condition, TIMSS has created numeric scales that weight answers from multiple questions about a topic. The following are the *Contextual Questionnaire Scales* in the 2011 TIMSS for grade 4 mathematics:

- Home Resources for Learning
- Early numeracy activities before beginning primary school
- Could do early numeracy tasks when began primary school
- Instruction affected by mathematics resource shortages
- Teacher working conditions
- School emphasis on academic success –principal reports
- Safe and orderly school
- School discipline and safety
- Students bullied at school
- Confidence in teaching mathematics
- Teacher career satisfaction scale
- Students like learning mathematics
- Students confident in mathematics
- Collaborate to improve teaching
- Instruction to engage students in learning
- Students engaged in mathematics lessons

Exhibit 3-5. Items in the TIMSS 2011 Early Numeracy Activities before Beginning Primary School Scale, Fourth Grade



Source: Methods and Procedures in TIMSS and PIRLS 2011. Available July 2013 online: <http://timssandpirls.bc.edu/methods/t-context-q-scales.html>.

To illustrate these scales, Exhibit 3-5 displays the items for the condition of early numeracy activities before beginning primary school scale in the fourth grade. The responses to these six items are pooled through an IRT statistical procedure to yield a scale with a mean across all countries of 10 and a standard deviation of 2. Cut points were then established on the scale to create three categories of doing early numeracy activities often, sometimes, and never or almost never (Exhibit 3-5 bottom).

International: OECD’s PISA (Program for International Student Assessment)

The Organization for Economic Co-operation and Development (OECD) launched the OECD Program for International Student Assessment (PISA) in 1997 to monitor the outcomes of education through measuring student achievement on a regular basis. PISA chose age 15 because “at this age students are approaching the end of compulsory education in most OECD countries.” PISA is known to focus on the use of knowledge in everyday tasks and challenges in reading, mathematics and science. PISA assessments are administered every three years, with each assessment emphasizing one of the three subjects, while assessing the other two subjects less extensively. The latest released assessment 2009 focused on reading. The 2009

assessment covered 67 countries and jurisdictions and, included for the first time the high performing jurisdictions of Singapore and Shanghai, China.

Along with the assessment of reading, mathematics and science literacy, PISA, like NAEP and TIMSS, includes a set of *contextual questionnaires* that ask students and the principals of their schools to respond to background questionnaires of around 30 minutes in length. Note that unlike TIMSS, PISA does not include a teacher questionnaire. For 2009 PISA, the questionnaires yielded information on:

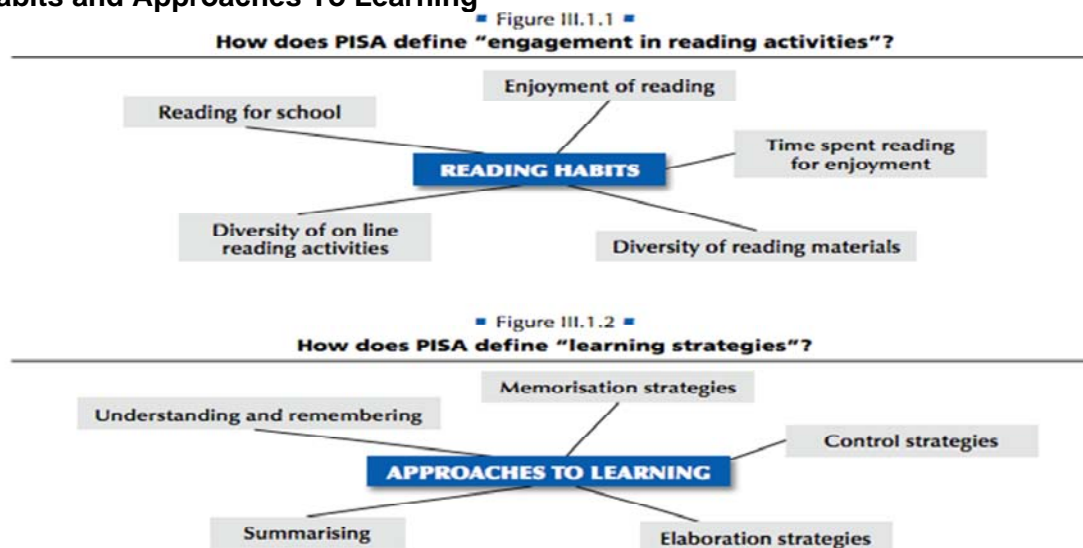
- *Students and their family backgrounds*, including their economic, social and cultural capital.
- *Aspects of students' lives*, such as their attitudes towards learning, their habits and life inside school, and their family environment.
- *Aspects of schools*, such as the quality of the schools' human and material resources, public and private control and funding, decision-making processes, staffing practices and the school's curricular emphasis and extra-curricular activities offered.
- *Context of instruction*, including institutional structures and types, class size, classroom and school climate and reading activities in class.
- *Aspects of learning and instruction* in reading, including students' interest, motivation and engagement.

The 2009 PISA, unlike NAEP, offered three optional questionnaires.

- *A computer familiarity questionnaire* focusing on the availability and use of information and communications technology (ICT), including where ICT is mostly used, as well as on the students' ability to carry out computer tasks and their attitudes towards computer use.
- *"An educational student career questionnaire* collecting additional information on interruptions of schooling and changes of schools, expected educational attainment and lessons or tutoring outside of school."
- *"A parent questionnaire* focusing on a number of topics including the student's past reading engagement, the parents' own reading engagement, home reading resources and support, and the parents' perceptions of and involvement in their child's school."

In the context of developing indicators PISA, like TIMSS but unlike NAEP, creates *scales* from multiple questions around a construct. But PISA goes beyond the TIMSS approach in studying the determinants of learning, by *estimating the quantitative relationship between the scales with learning*. This estimation process is illustrated for PISA scales for students' *reading habits*, as an indicator of engagement in reading activities *and approaches to learning*, as an indicator of learning strategies (Exhibit 3-6). Each of these scales is made up of five components drawn from student responses to the student questionnaire.

Exhibit 3-6. PISA Creating Numeric Scales From Multiple Questions About Reading Habits and Approaches To Learning



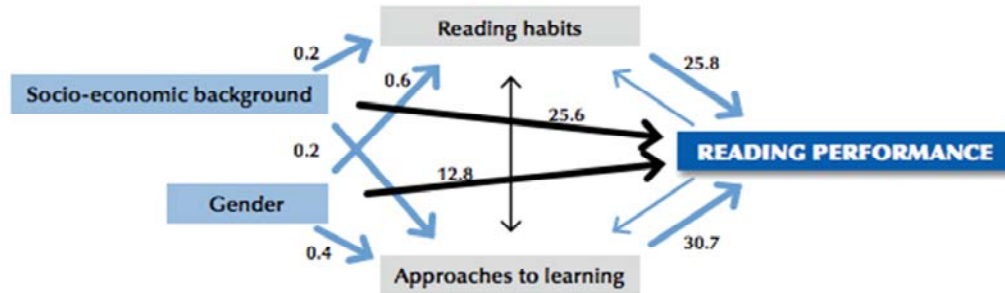
Source: OECD PISA 2009 Results: Learning to Learn

The two reading-explanatory scales in turn are employed in a path model (Exhibit 3-7) that relates students' socio-economic background and gender mediated through *reading habits* and *approaches to learning* to reading performance. For example, the difference that a unit change on the socio-economic background scale has on the reading performance TIMSS scale score consists of three effects. These are: a direct effect estimate of 25.6 reading score points, an indirect effect working through approaches to learning to reading performance of $.02 \times 30.7$ reading score points, and another indirect effect that comes from SES working through reading habits of 0.2×25.8 reading score points. The total effect then is $25.6 + (0.2 \times 30.7 = 6.4) + (0.2 \times 25.8 = 5.4) = 37.4$. One conclusion is that 68% ($25.6/37.4$) of the SES effect is direct.

Exhibit 3-7. PISA Statistical Estimation of Path Models Linking Socio-economic background and Gender Through Mediators With Reading Performance

■ Figure III.3.2 ■

How engagement in reading activities and learning strategies contribute to disparities in reading performance across OECD countries



Source: OECD PISA 2009 Results: Learning to Learn

The importance of particular indicator scales for learning is quantified in Exhibit 3-8, which shows the estimated net impact of socio-economic status if students were equally aware of effective reading approaches (i.e. summarizing strategies) as are students in the top quarter of the socio-economic scale. Thus, if students at the bottom quarter of the PISA index of economic, social and cultural status index had the same summarizing strategies for reading as students in the top quarter (Exhibit 3-8), they would diminish by one-third the total difference in reading levels between the bottom and third quartile of the socio-economic index.

Domestic: U.S. Department of Education's Priority Performance Goals

Under the Government Performance and Results Act (GPRA), the United States Department of Education (USED) is required to measure and report the performance of each of its Congressionally authorized programs. While many of the USED program specific performance indicators are not relevant for NAEP indicator-development, at least two areas are.

One is the *high-priority performance goals*. As part of this annual performance indicator activity, federal agencies are required to identify a sub-set of high-priority performance goals that are achievable within an 18 to 24-month time frame, with strong execution. The following are the *current USED K-12 high-priority indicators*:

- *Evidence Based Policy: Measuring Effectiveness and Investing in What Works: Implementation of a comprehensive approach to using evidence to inform the Department's policies and major initiatives, including:*
 - Increase by 2/3 the number of Department discretionary programs that use evaluation, performance measures and other program data for continuous improvement.
 - Implement rigorous evaluations for all of the Department's highest priority programs and initiatives.
 - Ensure all newly authorized Department discretionary programs include a rigorous evaluation component.
- *Struggling Schools Reform: National Models for School Reform: Identify as nationwide models 500 of the persistently lowest achieving schools initiating high-quality intensive reform efforts (e.g., turnarounds, restarts, transformations, or closures).*
- *Effective Teaching: World-Class Teaching and Learning: Improve the quality of teaching and learning by:*
 - Increasing by 200,000 the number of teachers for low income and minority students who are being recruited or retained to teach in hard-to-staff subjects and schools with rigorous, transparent and fair processes for determining teacher effectiveness
 - Ensuring that all States have in place comprehensive teacher evaluation systems, based on multiple measures of effectiveness including student growth, that may be used for professional development, retention, tenure, promotion, and compensation decisions.
- *Data Driven Decisions: Improved Achievement and Decision-Making through Statewide Data Systems: All States implementing comprehensive statewide longitudinal data systems that link student achievement and teacher data and link K-12 with higher education data and, to the extent possible, with*

pre-K and workforce data.

- *College and Career Ready Standards: World-Class College- and Career-Ready Standards: All States collaborating to develop and adopt internationally benchmarked college- and career-ready standards.*

While these identified high priority goals apply to federal programs, they, for the most part, deal with fundamental U.S. educational issues, ones that NAEP background questionnaire could inform. For example, NAEP can develop an indicator for data driven decisions. This would explore whether and how teachers use the data from their statewide longitudinal student outcome system to improve their practice. Principals could be queried whether they use the outcomes from the state longitudinal system to evaluate teachers. With respect to college and career ready standards, NAEP can go beyond adoption of these standards and track implementation challenges. In all three examples, NAEP can further ask about whether use of these measures has produced quantitative evidence of continuous improvement in the quality of teaching and the outcomes of students and how data use itself might be continually improved.

A second point of note is that USED has developed a set of *school-level leading indicators* for monitoring one of its new and controversial programs, School Improvement Grants (SIG) to support turning around low-performing schools. The idea of leading indicators is that they are precursors to turning around outcomes, in this case low-performing schools. NAEP indicators generated from its background variables might also be differentiated to single out those that are leading indicators of outcome changes based on research on student outcomes.

Domestic: Education Week's Quality Counts

Quality Counts is an annual report prepared by Education Week describing education performance indicators state-by-state. The state-by-state focus, unlike national indicators, allows for comparisons of education indicator values among states with similar student populations. A state can also compare its education indicator values to states with the highest education performance. As we shall discuss below, much of Quality Counts data on educational quality is based on state-level laws and requirements. NAEP, by obtaining data directly from principals, teachers, and students in the field, NAEP could provide unique complimentary information about how the provisions Quality Counts identifies are being implemented state-by-state and for many major urban districts.

Quality Counts annually reports on State education performance in six state education system areas. These are *performance indicators* and, unlike the previously discussed reports, Quality Counts gives each state *a letter grade* to gauge their performance in each of the six topic areas:

- *Chance for success* indicators looks at the connection between education and beneficial outcomes at each stage of a person's life. It covers indicators of education access, education outcomes and beneficial outcomes for early childhood, school years and adult educational outcomes.
- *School finance* indicators examine the level and equitable distribution of financial support within a state.
- *Transitions and alignment* indicators track state-policy efforts to coordinate the connections of K-12 schooling at three stages of education transition: early-childhood education, college readiness, and career readiness.
- *K-12 achievement* evaluates a state's student performance on three dimensions: current state performance, improvements over time, and equity as measured by poverty-based achievement gaps. The achievement indicators are all drawn from NAEP and complemented by rates for high school and advanced placement.
- *Standards, assessment and accountability* develops quality criteria for state implementation of these results-focused elements of state education policy.
- *The teaching profession* covers three aspects of state policy: accountability for teacher quality; incentives and allocation; and efforts to build and support the capacity of the teaching workforce.

Each of the six topic areas is composed of a number of components. For example, the standards indicator is composed of two components: states having course or grade-specific standards and providing supplementary resources or guides to implement the standards. The School Accountability indicator is composed of five components including school ratings, statewide student-identification system, rewards for high performing schools and assistance to low-performing schools, and sanctions for low-performing schools. Like the scales described above for TIMSS, Quality Counts creates an overall average, in this case an equal weighting of the scores given to each component and then assigns a grade.

The Quality Counts data are state level and for the most part do not report on actual implementation at the school and classroom level of the indicators. Quality Counts does annually supplement the State reports with an online field survey of registered users of the Education Week website with responses including views of teachers, instructional specialists, principals and other building administrators. While very informative, this annual survey examines a special topic, only, which in 2013 spotlighted school social and disciplinary environment and is not state representative. NAEP data could fill indicator gaps regarding implementation. For example, how schools and teachers are actually responding to the Common Core standards or how school finance differences might translate into real differences in

instruction, state-by-state and for students with different needs and backgrounds. NAEP can help provide these added data.

Domestic: NCES's Condition of Education

The Congress mandated that NCES produce an annual report on the *Condition of Education* to inform policymakers and the public about the current state and progress in key areas of education. The specifics of the report are left up to the Commissioner of Statistics. The latest 2013 report covers 42 indicators focused on four areas: population characteristics, participation in education, elementary and secondary education school characteristics and climate, and postsecondary education.

The following identifies the subset among the 42 indicators of most relevance for NAEP as those covering early childhood and elementary and secondary education.

Indicator Area 1. Population Characteristics

- Indicator 5, Percentage of children under 18 living in poverty (state and race ethnicity)

Indicator Area 2. Participation in education

- Spotlight on Preprimary Education: Kindergarten Entry Status: On-Time, Delayed-Entry, and Repeating Kindergartners
- Indicator 6. Enrollment Trends by Age
- Indicator 7. Early Education and Child Care Arrangements of Young Children (parents educational attainment)

Elementary/Secondary Enrollment

- Indicator 8. Public School Enrollment (state-level breakouts)
- Indicator 9. Charter School Enrollment (state-level breakouts)
- Indicator 10. Private School Enrollment
- Indicator 11. Racial/Ethnic Enrollment in Public Schools
- Indicator 12. English Language Learners (state-level breakouts)
- Indicator 13. Children and Youth With Disabilities

Indicator Area 3. Elementary and Secondary Education School Characteristics and Climate

- Indicator 16. Characteristics of Public Elementary and Secondary Schools (Urbanicity data)
- Indicator 17. Concentration of Public School Students Eligible for Free or Reduced-Price Lunch (school poverty, urbanicity breakouts)
- Indicator 18. Rates of School Crime
- Indicator 19. Teachers and Pupil/Teacher Ratios

Finance

- Indicator 20. Public School Revenue Sources (state-level breakouts)
- Indicator 21. Public School Expenditures
- Indicator 22. Education Expenditures by Country

Assessments

- Indicator 23. Reading Performance
- Indicator 24. Mathematics Performance
- Indicator 25. Reading and Mathematics Score Trend
- Indicator 26. International Assessments (states participating in international assessments breakouts)

Student Effort, Persistence, and Progress

- Indicator 27. High School Course-taking
- Indicator 28. Public High School Graduation Rates (state, race/ethnicity breakouts)
- Indicator 29. Status Dropout Rates (race/ethnicity breakouts)

Transition to College

- Indicator 30. Immediate Transition to College (family income breakouts)

Several points about these indicators are relevant for NAEP. One, the Condition of Education indicators cover the entire education system and as such NCES limits their numbers to the most strategic indicators. These strategic indicator areas for K-12 are focused primarily on student characteristics, finances and educational outcomes. However, for those interested in the quality of elementary and secondary education, NAEP surveys of principals, teachers and students can provide a great deal of information about school-level processes and students attitudes, learning out-of-school and use of time not contained in the Condition of Education report.

Two, the Condition of Education as a mandated report to Congress tends to have a national focus, with only a few of the indicators at the State level and no data for specific urban districts. NAEP's rich state-by-state and coverage of many urban districts can potentially add important disaggregated data to the Condition of Education content.

Third, NAEP has the potential to use its student assessment data to breakout Condition of Education indicators such as participation in education, school characteristics and climate and teacher characteristics by student proficiency levels.

Domestic: National Academy of Sciences' Key National Education Indicators

The National Academy of Sciences is in the process of establishing a list of key Education Indicators to be carried out by the Congressional Commission on Key

National Indicators. As part of their work in progress, they have published the results of a workshop on prospective frameworks along with a candidate list of key national indicators.

Exhibit 3- 9 National Academy of Sciences Draft Indicators Framework

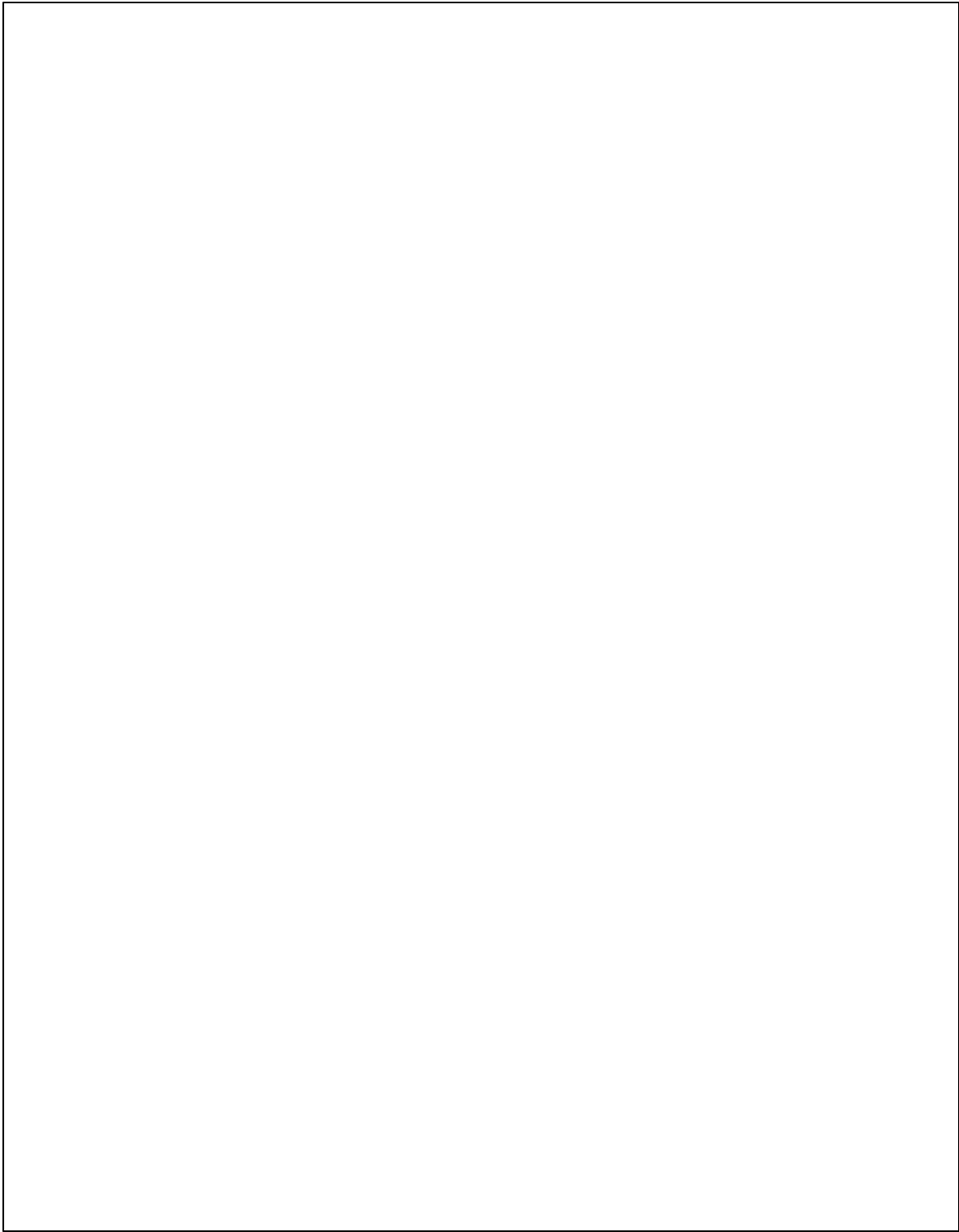
TABLE 1-1 Framework for Education Indicators Developed to Guide the Workshop

Stages of Education/Learning	Indicators About Institutions, Service Providers, and Resources	Indicators About Individual-Level Behaviors, Engagement, and Outcomes	Indicators About Contextual Factors
Birth to age 5			
K-12			
Higher education			
Other forms of postsecondary education and training			
Lifelong, informal learning			

Source. National Research Council, 2012

The draft framework emerging from the National Academy of Education Indicators Workshop (Exhibit 3-9) covers five stages of learning: preschool, K-12 education, higher education, other postsecondary education and training, and lifelong or informal learning (learning that occurs outside the formal structures of the education system). It also identifies three sectors of education: institutions, service providers, and resources; individual-level behaviors, engagement, and outcomes; and contextual factors that influence learning. This is similar to a production function process where column one combines inputs and processes, column two is outcomes of the education process; and column three is the interaction of the first two factors with the context in which education takes place.

The first two of the five stages in Exhibit 3-9 are most relevant for NAEP’s pre-college grade focus. The *indicators suggested for the preschool stage (Exhibit 3-10)* are organized according to the Exhibit 3- 9 framework into the providers and resources of education, individual student outcomes and contexts. At the preschool level, the institutions are the providers of early childhood education outside the home, the outcomes are both academic and social skills, and the context is the home environment including learning experiences from families and other caregivers. The NRC notes that the early childhood outcome measures could be through a NAEP assessment.



The indicators for K-12 education (Exhibit 3-11) also align with the Workshop

framework (Exhibit 3-9). These indicators include some common to the indicators above that report readily measured school services (teacher/pupil ratio) or student outcomes. But they also include a number of research-based school processes that require on-the-ground measurement from surveys or direct observations.



Examples of K-12 indicators proposed by the NAS workshop include:

- At the institution level include surveys of the safety and orderliness of the school climate and of the collaborative school community focused on student learning (using surveys of teachers, parents and students)
- At the teacher level of their mastery of content knowledge, quality of teacher-student interactions and proportion of teachers whose evaluations distinguish them from a basic standard, using measures of their contribution to student achievement and their professional practice.

NAEP surveys have the potential to provide address these rich in-depth information requirements.

Implications

The review of current indicator or indicator like reports yields a distillation of current practices that offer a range of possibilities for NAEP's indicator development design. Important practices derived from prior reports include:

- *Indicator frameworks* that include organization by a combination of age/grade range, production function-like models, policy issues or provider level in system.
- *Indicator selection* that may focus on regularly reported information, enacted policies or may instead describe services, processes and climates by in-depth school, teacher and classroom surveys.
- Development of indicators in the form of *indicator scales* that statistically combine results from multiple aspects about an indicator.
- The estimation of an indicator's importance in contributing to learning through *path models or other multivariate techniques* based on rigorous longitudinal research or more descriptive modeling based on cross-sectional data .
- The focus of indicator reporting at the *national level or at disaggregated sub-national state and major urban district* levels.
- Similarly, data can be reported nationally for all students or *disaggregated by student groups*.

The choices should be evidence based through direct and indirect relationships with key outcome measures

These factors along with traditional considerations of validity and reliability of indicator measures are considered in the following chapter on the choices and recommendations in designing and implementing a NAEP indicator framework.

4. Proposed Design of Key Indicators Framework

A NAEP indicators framework specifies five key features in designing a NAEP key indicators system: the organizing structure for the indicators framework; the specification of particular indicators; the development of indicator measures; survey and sampling considerations; and the reporting of indicator results.

Indicator Organizing Structure

The indicator organizational structure guides the selection of indicators by specifying the categories that sort out the selection of the most important factors affecting student learning. Further, the organizational structure clarifies the relationships among factors to learning, which aids in choosing measures and drawing implications.

Several indicator organizational structures were presented in the section 3 review of current domestic and international indicator and large-scale assessment systems. These offer different perspectives on the education system from which to choose potential components. At the outset, it should be noted that no one indicator organization is necessarily best, but it depends upon the purpose and focus of indicator systems. Looking across these indicator perspectives, and in the context of NAEP's focus on instruction, *five desirable design features* emerge in specifying an organizational structure applicable to the NAEP. These five features have been incorporated to form the proposed organizational structure for NAEP indicators in Exhibit 4-1, as follows:

1. *Explicitly including indicators that represent the key education drivers emerging in response to changing education or workplace conditions.* Indicators are most useful when they are used to monitor and continuously improve the education system to help achieve end outcomes and track responses in areas of major education change. Examples of drivers for the U.S. system might be preparing students with 21st century workplace skills, Common Core Standards, or instructional technology.

Focusing on major education drivers is consistent with the prior reports. Education At a Glance explicitly identifies policy issues to which the indicators relate. The NRC report begins its selection of K-12 indicators with a statement "that this system is the focus of many expectations, from producing responsible and productive citizens to boosting the nation's standing in science and technology and its position with respect to its economic competitors." EDWEEK explicitly builds tracking implementation of major policy reform areas, such as standards, assessment and accountability. Also note, that tracking education conditions in policy areas is not an endorsement of a policy approach but only a

consideration of the importance of monitoring and understanding responses to policy changes.

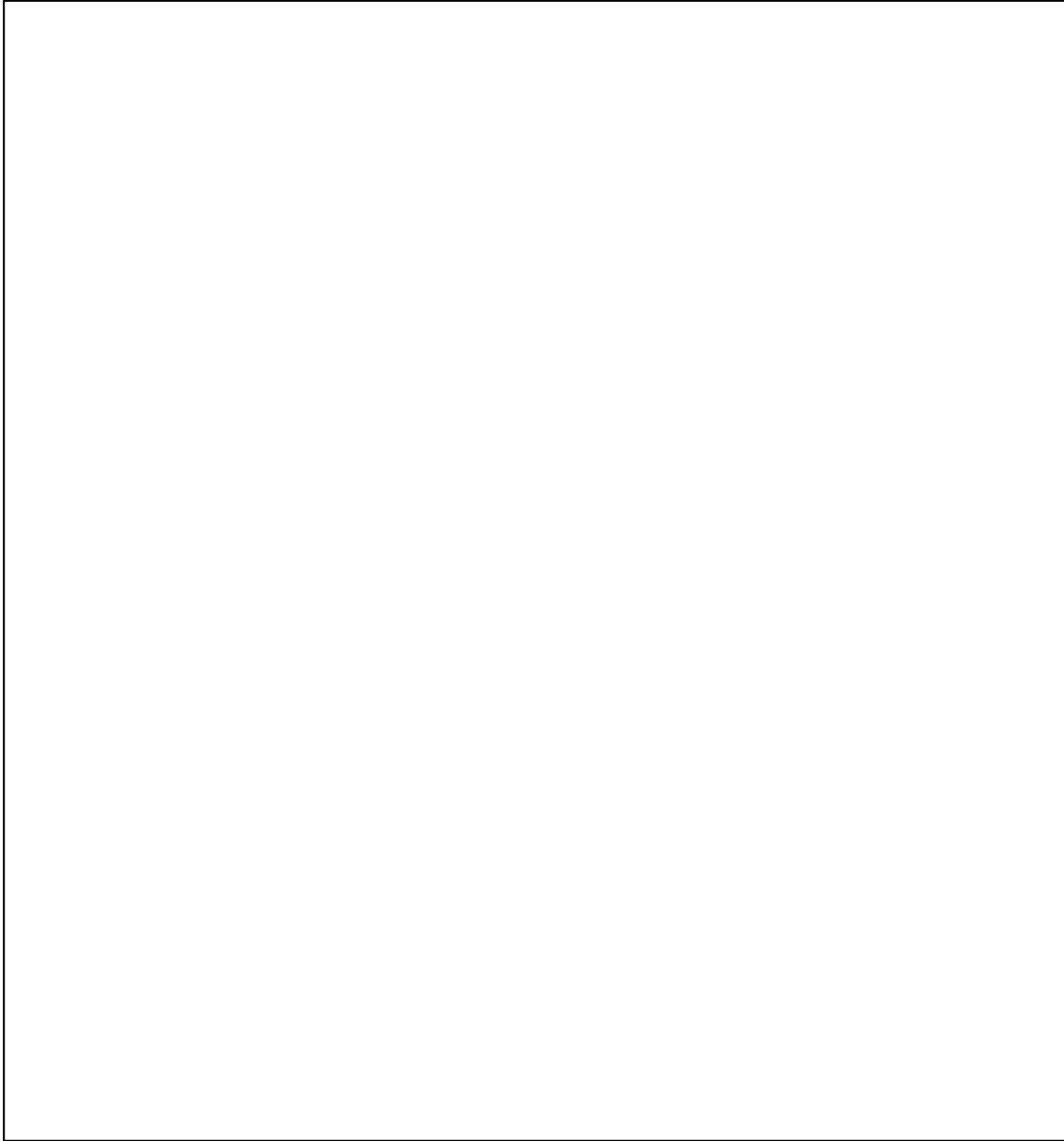
2. *Differentiating age/grade of instruction and learning.* The NRC report is explicitly organized around different stages of learning, from early childhood to adult and life-long learning. While NAEP is focused on K-12, the indicators framework should offer the possibility of including pre-school outcomes and experiences of students entering kindergarten. The framework may also want to differentiate early elementary, middle school and secondary school, as each has its own unique education aims and intervention priorities.

Exhibit 4- 1. Proposed Organizational Structure For an Indicators Framework, K-12			
A Specific Stage of Learning (Pre-Primary, Primary, Middle, Secondary)			
Locus of Education Activity	Key Drivers		
	Results	Enablers	Context/Constraints
Student			
Teacher			
School//Classroom			
System (district, state or nation)			

3. *Recognizing that the locus of education activity occurs at different levels at which education takes place – student, teacher, school/classroom and system – and develops indicators for each level.* Monitoring key education conditions translates into monitoring the key conditions at each of these levels. This is similar to the “Actors” identified in OECD’s Education At a Glance.
4. *Describing key education conditions in terms of education results (outcomes or outputs); the enablers which are the most important education factors producing education results; and the context and constraints within the education system that affects education results.* This focus on the broad elements in producing education is similar to the organizing structure used by Education At a Glance and the NRC analyses.
5. *Focusing on an indicator framework organization consistent with NAEP’s emphasis on instruction and learning as contributors to the NAEP assessment results.* This covers instruction and learning in both formal and informal settings. This emphasis on describing instruction and learning conditions is similar to the implicit indicators in TIMSS and PISA and to some extent the NRC proposed indicators. This focus differs from those of Education At a Glance or the Condition of Education, which tend to focus on the results of education rather than on instructional processes.

Indicator Selection

Indicator selection is the process of identifying key indicators that are essential to monitor on a regular basis. Parsimony in indicator selection is critical to prevent the indicator user from becoming overwhelmed in data with a loss of focus on priorities.



In selecting indicators for measurement, consideration should be given to:

- *Measuring what matters most.* Identifying what matters most should be guided by focusing on the education and context factors that research has shown bear an important relationship to results (Walberg, 2002). Variables that have high variance (e.g., differences in students' family socioeconomic

status) are ones that often have high contributions to outcomes.

Also, measuring what matters can give priority to informing policies. For instance, a widespread impression is that U.S. schools are having difficulty filling vacancies for mathematics teachers. Exhibit 4-2 drawn from TIMSS suggests that at least for U.S. middle schools, only about 12 percent of U.S. principals are having at least some difficulty filling vacancies for mathematics teachers. This compares with other Western English-speaking countries of 41 percent of the principals having difficulty hiring math teachers in Australia, 37 percent in England, and 44 percent in New Zealand. The United States numbers are slightly lower than those of Korea, but higher than Singapore. Of course, principal responses across countries may differ in the criteria they use for determining the adequacy of a candidate.

- *Focusing on changing conditions.* Continually measuring a factor that is an important contributor to results but changes infrequently produces little information gain from regular monitoring. This factor might be measured every other time NAEP is administered and the most recent value would be built into indicators that use multiple factors. On the other hand, education conditions undergoing rapid change such as because of technology or new policies should be weighted high for regular indicator development.

Taking a pyramid approach to indicator selection. A pyramid perspective on information recognizes that different users have different information needs. The top of the pyramid is the most important measure of a condition, such as national averages. Beneath this *top-level indicator*, additional indicator measures may display results disaggregated such as by population group, state, district or type of secondary school. The pyramid may display further information that shows indicator components, such as numbers, geometry, measurement, algebra and statistics for mathematics.

are drawn from the different international and domestic indicator lists above and represent an organized menu of indicator choices to guide selection of current and potential indicators for NAEP. Consistent with NAEP, the indicator structure is focused primarily around variables at student, teacher and school/classroom and system levels that support learning outcomes across the three aspects of education conditions (Exhibit 4-3):

- *Results* indicators include student assessment outcomes (such as from NAEP), but also teacher evaluations that include student outcomes, and other outcomes such as secondary school completion and parent satisfaction with the school.
- The *enablers* reflect formal learning from different levels of education. These include students exposure to preschool: teachers' knowledge and skills and their ability to apply them to create a challenging and supportive classroom learning environment: and school instructional time and student engagement in the content areas. Enablers also include system policies and regulations at district, state and national levels regarding teacher certification, standards, assessment and accountability.
- *Context/constraints* reflect factors not readily manipulable by the education systems, although conditions may be changeable with proper interventions, such as schools intervening in the home learning environment. These factors include learning at home and outside the school in formal and informal settings; factors influencing teacher quality including salaries and working conditions; and factors affecting the school learning environment including school safety, climate and class size.

Indicator Measurement

A sound measure for an indicator should meet criteria of validity, reliability, and consistency overtime.

Validity. A valid measure is one that adequately captures the underlying education condition of interest. Strong validity also depends on a good level of reliability. Occasionally a key indicator may be validly measured by a response to a single question, but more often a valid and robust indicator will be made up of multiple statistics each of which reflects an aspect of an education condition of interest. Combining multiple statistics such as responses from a number of questions around a topic into a larger comprehensive indicator measure or scale, is not an approach currently incorporated into NAEP background analyses.

Multiple questions with the same response stem can produce a scale based on response frequencies. We discussed how TIMSS grade 4 results for the early numeracy activities before beginning primary school are measured by responses to 6

questions about frequency of occurrence of these activities in terms of “often, sometimes or almost never” (Exhibit 3-5 above). The results in Exhibit 4-4 show how TIMSS creates a scale from these responses. For example, the scale for “often engaged in early numeracy activities” corresponds to parents responding to the six questions by indicating they do three of the six activities often and doing the other three sometimes. Within each country in Exhibit 4-4, students in families who on average across the six activities do these activities often score higher than students in families who sometimes do these six activities. They in turn score higher than students in families who never or almost never did these activities. However, these associations do not control for family background or other potentially important correlate factors.

Exhibit 4-4 Development of Indicator Scales From Multiple Questions

Source: IEA, *TIMSS*, 2011

Another example of creating a scale is used by Education Weeks Quality Counts is to give a letter grade based on a numeric score to each component forming an indicator and to average these scores to produce the letter grade. For example, the state standards, assessments and accountability indicator category is composed of the three subcategories. The subcategory for assessment consists of four assessment policies and is shown below:

EDWEEK Assessment Policies tracked

- **Types of Test Items:** For each item type, results are reported by school grade span. EPE Research Center review of testing calendars and other materials from state education agency Web sites, as verified by states, 2011.
- **Assessments Aligned to Standards:** Subjects in which state uses assessments aligned to state standards. Results are reported for each core academic-subject area. Ibid.
- **Vertically Equated Assessments:** State tests for the 2011-12 school year have been vertically equated in grades 3-8 so that scores for each grade have been placed on a common metric. Results are reported for English/language arts and mathematics. EPE Research Center annual state policy survey, 2011.
- **Benchmark Assessments:** State provides educators with benchmark assessments or item banks linked to state standards. Assessments or test items may be developed by the state or an external organization.

This subcategory assessment is scored “reflecting the percent of tracked policies a

Exhibit 4-5. Quality Counts Use of Average Indicator Scales

Maryland		Overall: B+ (87.8)
Chance for Success: B+ (87.8)	The Teaching Profession: B (83.7)	
Early foundations: A- (90.8)	Accountability for quality: C (76.5)	
School years: B (84.2)	Incentives & allocation: B (84.6)	
Adult outcomes: A- (91.1)	Building & supporting capacity: A- (90.0)	
K-12 Achievement: B (83.9)	School Finance: B+ (86.6)	
Status: B (85.4)	Equity: B+ (87.3)	
Change: B+ (89.3)	Spending: B (85.8)	
Equity: C- (71.7)		
Standards, Assessments, & Accountability: B+ (88.3)	Transitions & Alignment: A (96.4)	
Standards: A (96.4)	Early-childhood education: A (100.0)	
Assessments: C+ (78.3)	College readiness: A- (90.0)	
School accountability: A- (90.0)	Economy & workforce: A (100.0)	
Want more data? Get the State Highlights Reports >>		

Source: EDWeek, Quality Counts, 2012

state has implemented” and a numeric score is assigned the subcategory. The scores are then averaged across the state standards, assessment and accountability subcategory and letter grades are assigned based on scores (A=93 to 100, A-minus =90-92). Exhibit 4-5 shows a full Quality Counts display for Maryland, the highest rated state by 2012 Quality Counts.

Reliability. A reliable measure is one where the indicator measure produces consistent results when repeatedly measuring the same underlying condition. Of

particular note for measures based on surveys of background variables is that they often use qualitative responses to questions about frequency of occurrence of an activity. For instance, a question might ask for a response in terms of “a lot” or “a little” that are subject to interpretation and may be sensitive to respondent context. For example, in Exhibit 3.5 shows how TIMSS asks families about the frequency of

Exhibit 4-6 Differences between teacher and school reported responses about science resource availability raise issues of response reliability

	Not at all	Small Extent	Moderate extent	Large extent
Science Kits are provided (teacher reported)	26	30	29	16
Science Kits are provided (school reported)	7	24	32	37
Science magazines and books are provided (teacher reported)	22	35	33	11
Science magazines and books are provided (school reported)	2	19	35	44

NAEP Data Explorer

early numeracy activities. A parent with only a high school education or less may interpret “a lot” different than for a parent with a college degree.

Qualitative responses may also be sensitive to the respondent. In the recently completed NAEP background paper on science Exhibit 4-6 was presented showing that teachers were more likely to indicate that resources within a school were “not at all available” than were principals in the same school. This is not surprising, as it is principals who are responsible for school resource availability. Conversely, a strong indicator of a positive school climate might be the degree to which principals and teachers agree on school quality factors including availability of science resources.

Consistency. A consistent measure requires using the same measure for an indicator over time. To the extent that measures are changed from time period to time period then it is unclear whether a change in an indicator condition comes about because of a real change in the underlying condition or because of changes in the measure. The Expert Panel report addressed this issue in recommendation 1d:

“Use consistency over time as a criterion to consider for question selection and wording. NAEP’s inconsistent inclusion of background questions weakens its potential to track trends and improvements within a subject area and topic.

For example, the Expert Panel found that only one-third of the 2011 questions asking about course offerings yielded at least a 6-year trend. No 2011 questions about curriculum or school resources were found on the 2005 or earlier questionnaires.

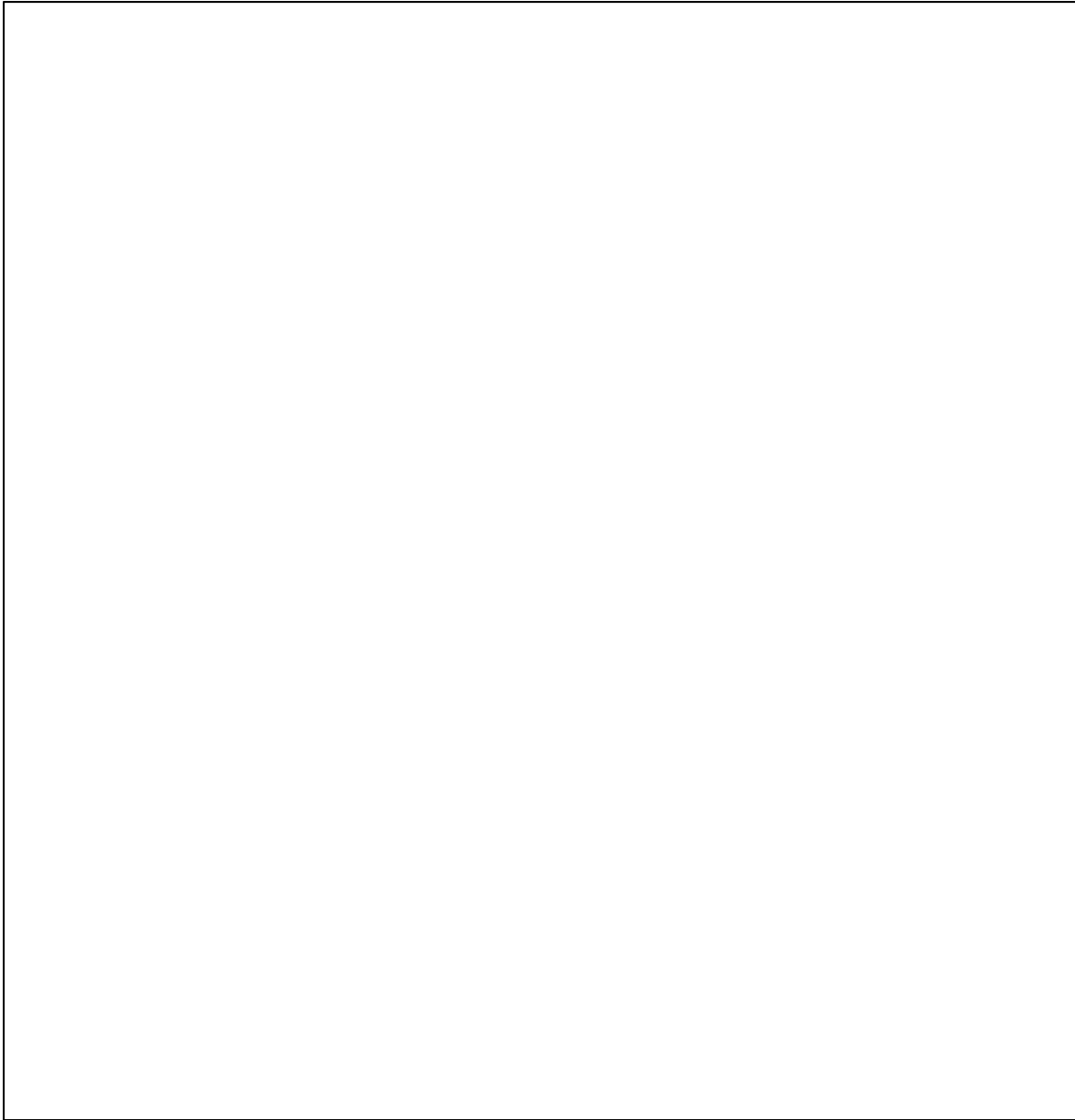
Sources of Indicator Data

Many of the indicator measures across the international and domestic reports are derived from surveys generated during the report process. A second source of indicator data draws upon secondary sources from other surveys. Both are considerations in specifying an indicator framework.

With respect to data from surveys, TIMSS and PISA, unlike NAEP, incorporate a household survey to directly obtain information about parents or guardians socio-economic status and about the home learning environment. TIMSS innovatively combined with PIRLS to develop a joint household survey for grade 4 students. The household survey included questions about:

- Early numeracy activities in the home before beginning primary school (See Exhibit 3-5)
- Early literacy activities in the home before beginning primary school
- Amount of exposure to preschool
- Family perception about child's literacy and numeracy skills before entering primary school
- Family interaction with the child about school work
- Family perceptions about school
- Family literacy environment
- Family SES

These represent an extensive set of questions about student and family home learning and socio-economic environment conditions compared with NAEP, with its only source of grade 4 information derived from a brief grade 4 student questionnaire. As an example, Exhibit 4-7 displays the results from the home responses on how well their children could do when entering primary school on six numeracy tasks. In every country, average grade 4 mathematics achievement declined as parents reported that their entering primary children could do fewer tasks. This correlation lends external validation to parent responses.



A second characteristic of several of the indicator reports is the pooling of information across different surveys. The Condition of Education and Education At a Glance are drawn almost entirely from data series generated by from other surveys. Quality Counts is also a state-level amalgam of Education Week's direct analyses of state policies combined with data from other surveys, including prominently featuring the NAEP assessment results. Currently, NAEP background variables only include those from the NAEP student, teacher and school surveys, but combining NAEP background data with data reported from other surveys is a potential source of expanded background reporting.

Reporting on Indicators

A challenge in reporting on indicators is that different audiences need different levels of depth of indicator reporting. In response, the business sector and more recently government have implemented *digital dash boards*. These provide a click-of-the-mouse approach to presenting different visual perspectives on key performance indicators.

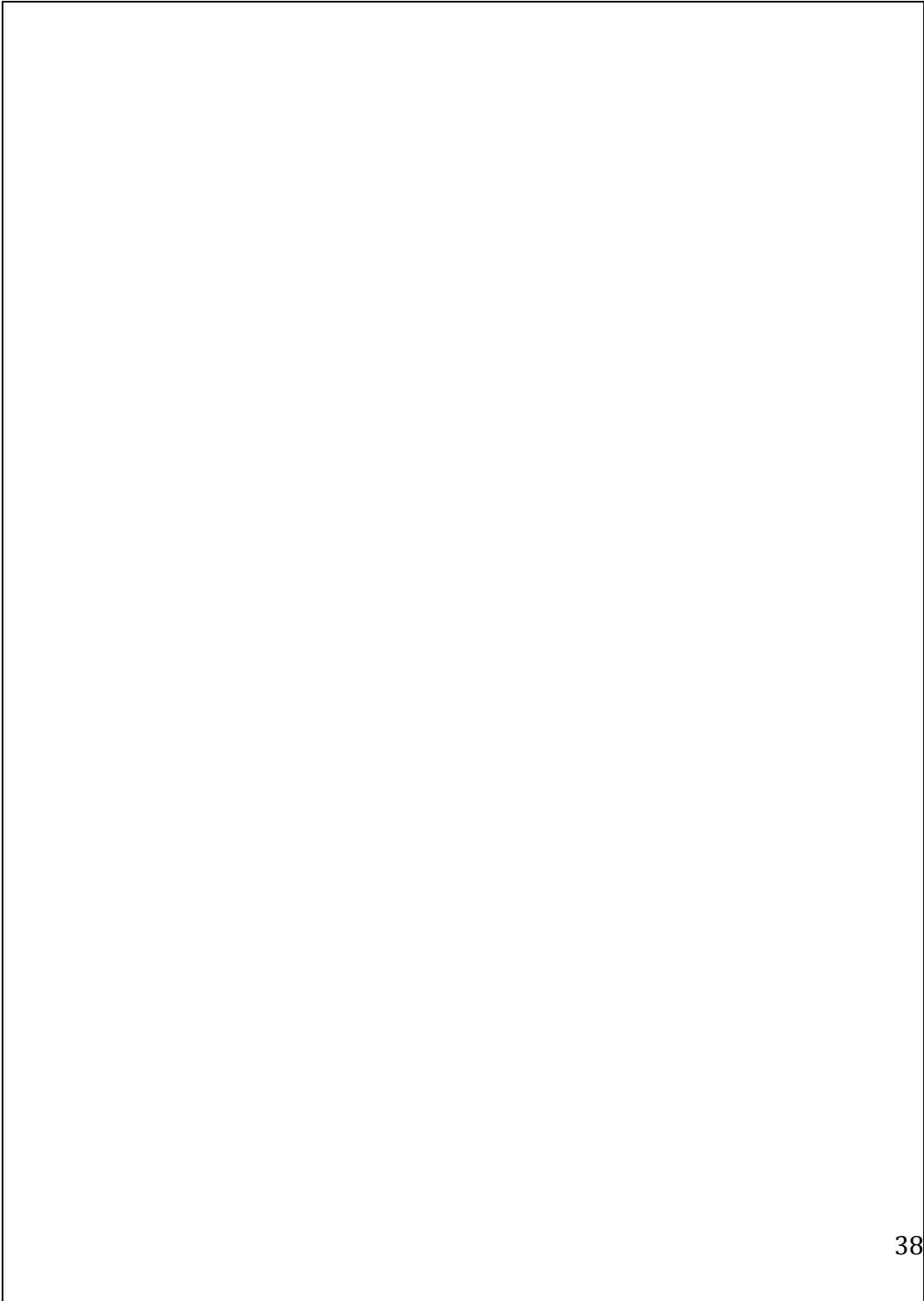


Exhibit 4-8 illustrates the dashboard presented to web site users to the U.S. Department of Education's performance indicators for teachers and leaders. There are three indicators with national summary results and direction of change shown for each. From the dashboard the interested user can drill down and get a chart of the data, state comparisons and details on an indicator including its specific goal statement, how it is measured, why it is important, etc.

5. Next Steps: Using the International and Domestic Indicator Framework to Guide Development of a NAEP Indicator Framework and Provide Examples With Current Data

A follow-on to this report will be a second indicator report to NAGB by December 2013 that will contain *a recommended set of Key Indicators, examples using current NAEP data, and recommended improvements in NAEP data to strengthen indicator measurement or fill indicator gaps*. The second report will build off of the findings in the initial report by addressing the following topics:

- Specify a NAEP Indicators Framework for Background Variables applicable across cognitive assessments.
- Identify indicators that are estimable using current NAEP data (and present a set of examples); indicators that could be developed through changes in the NAEP questionnaires; and indicators that would require a fundamentally new NAEP questionnaire.
- Identify where NAEP offers a unique data or measurement advantage over other indicator sources.
- Explore opportunities for combining NAEP with other NCES indicator-supporting data.
- Explore opportunities for aligning NAEP domestic indicators with the indicators generated by background variables in the international data collections discussed above to yield national and state comparisons with other countries.
- Explore how can NAEP reports best display a pyramid information approach along the lines of an indicator dashboard to provide the user with push-button access to top-level national measures or to more disaggregated measures by indicator component, student characteristics or jurisdictions.

- Assess how consistently the identified key NAEP education indicators have been measured by NAEP over time and identify challenges in fixing these definitions.

References

- Aud, S., Wilkinson-Flicker, S., Kristapovich, P., Rathbun, A., Wang, X., and Zhang, J. (2013). *The condition of education 2013* (NCES 2013-037). U.S. Department of Education, National Center for Education Statistics. Washington, DC. Retrieved [date] from <http://nces.ed.gov/pubsearch>.
- Ginsburg, A., Noell, J. and Plisko, V. (1988). Lessons learned from the wall chart. *Educational Evaluation and Policy Analysis*, 10(1), 1-12.
- Government Accountability Office (2011). *Key indicator systems: experiences of other national and subnational systems offer insights for the united states*. Available July 2013 online: <http://www.gao.gov/assets/320/317346.pdf>.
- Government Performance and Results Act of 1993, Pub. L. No. 103-62, 107 Stat. 285 (1993).
- Education Week (2013). *Quality Counts*. Available July 2013 online: <http://www.edweek.org/ew/toc/2013/01/10/index.html?intc=EW-QC13-LFTNAV>
- International Association for the Evaluation of Educational Achievement (IEA) (2012). *PIRLs 2011 International Results in Reading* Ina V.S. Mullis, Michael O. Martin, Pierre Foy, and Kathleen T. Drucker
- International Association for the Evaluation of Educational Achievement (IEA) (2012). *TIMSS 2011 International Results in Mathematics* Ina V.S. Mullis, Michael O. Martin, Pierre Foy, and Alka Arora.
- National Research Council. (2012). *Key National Education Indicators: Workshop Summary*. Steering Committee on Workshop on Key National Education Indicators, A. Beatty and J.A. Koenig, Rapporteurs. Board on Testing and Assessment and Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- OECD (2010), *PISA 2009 results: learning to learn – student engagement, strategies and practices (volume III)*. Available July 2013 online: <http://dx.doi.org/10.1787/9789264083943-en>
- OECD (2013), *Education at a glance 2013: OECD indicators*, OECD Publishing. <http://dx.doi.org/10.1787/eag-2013-en>
- Smith, M., Chudowsky, N., Ginsburg, A., Hauser, R., Jennings, J. and Lewis, S. (2012). *NAEP background questions: An underused national resource*. A report to the National Assessment Governing Board by the Expert Panel on Strengthening the NAEP Background Questions. National Assessment Governing Board. Available online at: <http://www.nagb.org/content/nagb/assets/documents/publications/expert-panel-naep-bq-report.pdf>
- Special Study Panel on Education Indicators. (1991, September). *Education counts: An indicator system to monitor the nation's educational health*. A report to the Acting Commissioner of Education Statistics.

U.S. Department of Education (2012). *The department's FY 2012-13 priority performance goals*. Available online at:
<http://www2.ed.gov/about/overview/focus/goals.html>.

Walberg, H. (2002). *The national assessment's most useful background items*. National Assessment Governing Board.