Creating Assessment Frameworks: Experience from International Studies

Larry E. Suter
Educational Research Consultant

May 15, 2012

Introduction:

Cross-national assessments of student learning in mathematics, science, reading, computer technology, and civics have been successfully conducted since the 1960's. Each subject required professional researchers and educators from different cultural backgrounds to reach agreement on a common definition of the content areas and measurement techniques for formal schooling. Two international organizations, the International Association for the Evaluation of Educational Achievement (IEA) and the Organization for Economic and Cooperation and Development (OECD) are now continuously conducting successful cross-national surveys of science. Could these studies provide useful models for establishing new assessments of science in informal learning environments? I began answering this question about whether the conceptual frameworks designed for the international studies could provide a basis for creating new assessments of informal learning. Second, could the processes of establishing agreements among educators from different cultural backgrounds about the topics to be assessed in formal education also be applied to creating an agreement on what to assess in the diverse forms of informal learning? Finally, I wondered whether nationally representative descriptions of the activities to be assessed were a necessary ingredient for creating conceptual frameworks. This paper presents introductory answers to these questions.

Definition of a Framework for Educational Assessment

Educational assessments require a common definition of the domains of cognitive knowledge. They are based on professional judgment about the scope of an area of content knowledge; thus the content of frameworks may differ from place to place and time to time. A Framework provides the definition of the elements of the assessment and is based on a theoretical understanding of the development of human learning of the topic that can guide the development of test items and scales. The latest NRC report defines their framework as, “a broad description of the content and sequence of learning expected of all students by the completion of high school” (NRC 2012). The authors of the NRC science framework point out that a framework need not contain all specific details of science content but contains a general guide of where to look. One characteristic of assessment frameworks is that the content usually has been adopted by a community of recognized practitioners and scholars, professional societies (such as National Council for the Teachers of Mathematics (NCTM), the National Research Council (NRC), or the OECD governing board. Appendix A contains a brief summary of some of the content in selected assessment frameworks in science and mathematics from formal education and informal education. They differ in specifics of content in the theory of learning underlying the frameworks but agree in general content and structure of science.

Assessment frameworks usually have multiple dimensions made of science content in one dimension and categories of knowledge, such as the Bloom’s taxonomy of learning (Appendix A), as another dimension. However, as data are gathered and analyzed in repeated surveys the frameworks themselves are often modified (Robitaille 1996). Reasons for altering the frameworks include concerns about how well the statistical results had informed practices of creating curricula in the subject areas (Mullis, 2011) or with the application of theories of cognitive behavior of individuals. For example, the original framework for the Third International Mathematics and Science Study (TIMSS) was modified from earlier international frameworks for the Second International Mathematics Study (SIMS) as the relationship between a specific content area and cognitive behaviors were modified by mathematics educators (Robitaille, 1992, p 42 and Mullis, 2011). For a discussion of the reasons for altering National Assessment of Education Progress (NAEP) frameworks and specifications in 2009 and for an explanation of the types of changes that were made see the reports on the National Assessment Governing Board (NAGB) website (NAGB 2007). The opening discussion of the 2009 science item specifications gives four reasons for altering the frameworks: National standards were published for science; scientific research itself had advanced, research on cognitive learning processes had advanced, and assessments were gaining in national and international prevalence in education (NAGB 2007). Thus, frameworks for assessment are always a in progress of change depending on the discoveries of science and the practical uses to which the assessments are applied.

The two well-known international assessments of science are The Third International Mathematics and Science Study (TIMSS) and the Program for International Student Assessment (PISA). Each study has different origins and different resulting frameworks for science assessment.

The TIMSS framework is school based with a theoretical structure that seeks to identify elements of the content of instruction in the intended curriculum and relate the impact of that curriculum to instruction (implemented curriculum) and student achievement (attained curriculum). The basic elements of the framework were developed from early attempts in the 1950's and 1960's to "measure the unmeasurable" (Husén 1975) across countries. These frameworks were developed from topics found in textbooks and national definitions of content areas, not from general theory of learning (Dossey communication). The frameworks provided an outline of the topics that would be measured by specific items that could be administered in a student test. New test items were created to match the framework and the items have to be tested for proper psychometric properties. Consequently, not all topics in the framework were able to be fit with sufficient items. As a result of testing and summarizing the results of surveys, the frameworks were modified over a period of 40 years of publication and efforts to apply the results to practice (Robitaille 1992 and Mullis 2011).

The PISA framework began through a cooperative system of efforts by OECD to establish a new basis for an international educational statistical system that was parallel to the existing statistical system on employment. The PISA framework explicitly includes "inquiry" as an aspect of science learning (Appendix A). International committees of education and science experts were established and they chose to assess the competencies that are used as standards of good work in Europe. The test items tend to be items of general problems faced by engineers, for example, but that could be solved with logic and experience (rather than rote memory). However, the analyses of these data tended to draw attention to the school system rather than the employment system for explanations of the country differences; thus the frameworks have been modified to include a greater attention to the content of instruction. As the surveys were published and used for policy direction, the survey directors found that they needed to modify the original frameworks. A continued series of data collection and analysis is a necessary element of conceptual development of assessment frameworks.

The content frameworks provide a structure for item selection and scale development for a student assessment. For example, the content framework for TIMSS science lists 80 topics in science areas that are likely to be found in elementary-secondary school classrooms. PISA created a framework that defined general properties of education without specific content areas. Hundreds of specific test items are written by item writers for the key topics of the framework.

The final test length does not permit complete coverage of each topic; therefore items are selected for the final test to represent a sample of the content areas. The final set of items chosen for the assessment must have statistical properties that permit generalization and representation of an underlying construct (mathematics, or algebra). Scales are created from the chosen items with statistical procures such as the IRT model that force items to be ranked from high to low ability. The results of these scales are then evaluated by content experts to judge whether they actually appear to reflect the content that was intended in the framework. In recent years, multiple scales have been developed to represent different aspects of the content areas; however, the number of scales that can be created is limited by the number of items used in the survey; consequently, the representation of science content is usually summarized.

Each of the international studies was monitored by prestigious oversight committees that reported to the funding agencies. The IEA international studies were monitored between 1987 and 2000 by the NAS Board on International Comparative Studies (BICSE) which met regularly and discussed issues of assessment and statistical comparisons with formative responsibility for the development of the studies. The PISA evolved from an international committee structure developed in 1989 to create improved communication of purposes of international comparisons across countries. The OECD governing board, made up of official country representatives, continues such oversight of the study collection and analysis. These committees had a large influence on the content and quality of the studies.

Non-Cognitive assessment

The two international frameworks contain a set of categories for defining cognitive behaviors such as knowing, applying, and reasoning (Appendix A). But the frameworks for science learning rarely include references to specific attitudes or engagement practices. The 1995 TIMSS framework contained contain a dimension called "Perspectives" that was made up of attitudes, career expectations, and habits of mind. These were chosen for assessment because they are characteristics of persons who continue with a career in the science fields (Robitaille 1992). But that dimension was never fully populated with test items and was dropped from TIMSS frameworks after 2000 (Mullis 2011). While items on attitudes of students toward science in their lives were included in both TIMSS and PISA surveys these items were not integrated into the assessment frameworks of science knowledge. They are conceived mainly as explanatory concepts rather than outcomes of science experiences. The TIMSS attitude items do not appear to emanate from an existing theory of learning or goal setting. The PISA surveys include an extensive battery of attitude scales (see Appendix B) that are related to country level performances in publications of science achievement. The 2006 PISA study experimented with models and collection methods that integrate attitude toward scientific investigations with measures of cognitive knowledge (OECD 2006, chapter 3). They report that students were asked attitude questions about three of the cognitive items to measure whether the student supported the notion of scientific enquiry in the topic of the question. They report that in some cases, students strongly support conducting research on a topic (such as viruses) but that they were not confident that scientific investigations would provide sufficient answers. In general, they found that the association between scales of support for scientific understanding and test scores themselves was positively related, but the average level of association is weak at .20 (PISA 2006 Chapter 3). Thus, it is possible to conclude that the two dimensions – cognitive and attitudinal - are somewhat independent and worthy of separate measurement. Each attitude scale of beliefs, values, goals, interests, and efficacy (Appendix B) could provide useful information to guide policy making and practice about scientific investigation.

Summary and Recommendations:

Lessons learned from this review of establishing frameworks for student assessment are:

1. The international studies frameworks attempted to identify the topics studied in school (TIMSS) and those competencies that were necessary for successful work (PISA). The content areas were established through <u>continuing interaction</u> between formal theories of learning, pre-existing standards (such as textbooks), and data (results of small scale surveys).
2. The international studies <u>established a demand for measurement</u> from funding agencies by reporting information of use to policy makers and the general public. Because of this demand, many researchers were willing to join in the formative stage of developing the survey instruments.
3. The international studies <u>borrowed theories</u> for assessment from cognitive sciences, communication theory, motivation theories, and evaluation practices of informal learning. The results of the studies also provided new insights into the theories themselves. Projects funded by NSF, the National Center for Education Statistics (NCES), and other foundations supported the continuing development of theories of informal learning.
4. The international studies <u>created intellectual debate</u> by making the final resolution of the content of the framework the responsibility of several individuals working separately to make forthright interpretations and recommendations. The publication of results was

followed with a debate of the framework in ongoing sessions at NAS through an ongoing national committee to create a framework.

5. The creators of frameworks included participants from content areas (scientists) and practitioners (educators) who contribute to the definition of the changing nature of science.
6. The content of assessment frameworks are mostly about the cognitive dimensions of science rather than affective or conative. However, the international data collections obtained separate measurements of goals, values, and interests toward science. The original framework for TIMSS included a dimension of attitudes, habits of mind, and career interests that was never populated with test items.

Recommendations:

1) The development of a general framework for informal science assessment is needed to guide the development of item specifications of the necessary attitudes, goals, and engagements that are known to be affected by attending informal science settings (see NAGB 2007 for an example).
2) Establishing a consensus framework on informal science outcomes will not be easily accomplished. While this conference will stimulate greater interest in and intellectual basis for new assessments of informal activities, the evidence from previous attempts to measure adult understanding of science is not encouraging. For example, thoughtful professional educators themselves do not agree on the outcomes of informal learning (for example, see Rychen and Salganik 2001). Creating a formal definition of categories raises the fear that, once a subject has been formally defined, all individuality may be lost. Data and conceptual

frameworks are believed to get in the way of practice and existing policy (Rychen 2001). These same fears were found in the 1990's when the TIMSS framework was first developed and was presented as an alternative to existing assessments of factual knowledge (although in fact the final product was not significantly different). Moreover, attempts have been made and debated through such places as science conferences, journals such as the *Public Understanding of Science,* in the National Science Board's *Science Indicators*, and surveys of adult learning. Yet no common framework has been created because of the rapidly changing nature of science itself and of public understanding (National Science Board 2012, page 7-20).

3) Researchers interested in informal learning should expand the variety and forms of collections of national data to guide conceptualization. Well conducted studies are needed to demonstrate whether informal activities are consequential in the lives of individuals. A growing federal interest in accountability of federal funding programs in informal learning could be addressed by such studies.
4) A new international study of informal science learning would provide a greater stimulus to develop the framework and create corresponding assessment items. It would add to the discussion of the outcomes of visiting informal institutions. International comparative studies of informal science learning and engagement would be likely to be well received by publishers and funding agencies.
5) The existing research on non-cognitive measures of attitudes toward science (as conducted by PISA and many smaller US national studies) should be improved, standardized with better theory, and related to cognition. Informal science experiences in museums, aquaria, after-school programs, and science television programs create opportunities for exploration,

curiosity, and experimentation with role models of scientific engagement. Research should be supported to conduct further studies of measuring the role of informal settings in maintaining or producing curiosity of science.

6) Better descriptions of informal science activities are needed to support the basic elements of a framework. Unlike formal schooling, informal learning practices are not institutionalized through curriculum, textbooks or standards that can be described and evaluated. New more detailed surveys of students and adults that describe the entire landscape of US children and adult participation in informal settings would help establish the possible goals of informal learning worthy of assessment (and rule some out). Measuring learning of topics that are known to occur among children and adults is better than measuring an ideal that is infrequent. Some national surveys already exist that provide some insights into the proportion of the population engaging in informal science settings and provide an opportunity for further analysis and distribution to policy makers (see Appendix C) .

7) An informal assessment framework should be consistent with the latest theories of human learning and on knowledge of maintaining engagement, inquiry, and curiosity with science. The framework should define relevant attitudes, goals, and values toward science that are perpetuated by attending informal education centers rather than be restricted to cognitive knowledge. These topics are also found in the first Dimension of Science of 2012 NAS Science Framework (2012). An informal science assessment framework might use the 1992 TIMSS framework category "perspectives" that includes: attitudes toward science, career expectations, habits-of-mind, reasoning, and communicating.

8) Informal assessments will need to create methods of measuring achievement that do not require responses to formal exams because informal learning experiences seldom allow the

participants to take tests or respond to serious questions. The use of computer technology in informal settings creates new opportunities for data collection during informal experience.

9) Informal science assessments could borrow general frameworks from adult science literacy. For example, a set of categories proposed by researchers for the NSF Indicators biannual report for measuring adult learning of science include the topics of civic engagement with science, practical decision making, and curiosity about scientific world view. The framework might include a cognitive dimension of factual knowledge, knowledge of processes in science, and knowledge of science institutions.

Conclusion:

The international comparative studies provide some useful examples of how to achieve agreements on a definition of the content of human learning. However, the experience demonstrates that agreement is not immediate and the frameworks themselves are altered with experience in using the resulting data bases. A national framework for informal learning in science could be developed from aspects of existing international assessments of formal learning; especially those aspects that focus on non-cognitive behaviors. Informal learning is intentionally based more on living experiences and participation in scientific processes rather than memorization of already-digested content knowledge. An assessment framework needs the assistance of empirical evidence of the scope of practices in informal settings already underway. Conducting an international accounting of after school activities would be a dynamic way to obtain cooperation and collaboration from a wide audience of practitioners, scientists, educational leaders, and policy organizations.

Appendix A.

Science Assessment Frameworks Content

I. *Achieve*:

Content performance

- Biology
- Chemistry
- Physics
- Earth/Space Science

Cross-Cutting and Interdisciplinary Content

- Cognitive Demand
- Knowing
- Applying
- Reasoning
- Inquiry Skills
  - Basic
  - Advanced

II. International assessments of science and mathematics (year of collection)

1. First International Mathematics Study (1964)
2. Second International Mathematics Study (1982)
3. Second international Science Study: (1987)
4. Third International Mathematics and Science Study: (1995)

   - Content areas (specific scientific disciplines)
   - Performance Expectations (Understanding, theorizing, using tools, investigating, communicating)
   - Perspectives (attitudes, career expectations, participation, interest, safety, and habits of mind)

*III.*Trends in International Mathematics and Science Study: 2011

Content (scientific areas)
Cognitive domains (knowing, applying, and reasoning)

5. International Adult Literacy Study (2001)
6. International Computer Literacy Study (1990)
7. Program in International Assessment (2000-2009)
   - A. Reading
   - B. Math
   - C. Science

- <u>Context</u>: recognizing life situations involving science and technology.
- <u>Knowledge:</u> Understanding the natural world on the basis of scientific knowledge that includes both knowledge of the natural world, and knowledge about science itself.
- <u>Competencie</u>s: demonstrating scientific competencies that include identifying scientific issues, explaining phenomena scientifically, and drawing conclusions based on evidence.
- <u>Attitudes</u>: indicating an interest in science, support for scientific enquiry, and motivation to act responsibly towards, for example, natural resources and environments.

IV.<u>NSF Informal Science Evaluation Framework:</u>

- *Awareness, knowledge, or understanding* of STEM concepts, processes, or careers.
- *Engagement or interest* in STEM, concepts, processes, or careers.
- *Attitude* towards STEM-related topic or capabilities.
- *Behavior* resulting from engagement.
- *New skills* based on engagement

V.<u>Public Understanding of Science Frameworks:</u>

- civic engagement with science,
- practical decision making, and
- curiosity about scientific worldview

By three content areas:

- factual knowledge,
- process and standards, and
- Institutional knowledge.

VI.<u>Bloom's Taxonomy</u>

A. *Factual Knowledge* - The basic elements that students must know to be acquainted with a discipline or solve problems in it. Terminology, details and elements.

B. *Conceptual Knowledge* - The interrelationships among the basic elements within a larger structure that enable them to function together. Classifications, categories, principles, generalizations, theories, models, and structures.

C. *Procedural Knowledge* - How to do something; methods of inquiry, and criteria for using skills, algorithms, techniques, and methods.

D. Metacognitive Knowledge - Knowledge of cognition in general as well as awareness and knowledge of one's own cognition: Strategic, contextual, and conditional knowledge and Self-knowledge.

Appendix B

PISA 2009 Attitude Scales

1. General value of science
    a. Science is important for helping us to understand the natural world.
    b. Advances in science and technology usually improve people's living conditions.
    c. Science is valuable to society.
    d. Advances in science and technology usually help to improve the economy.
    e. Advances in science and technology usually bring social benefits.
2. **Index of personal value of science**
    a. I find that science helps me to understand things around me.
    b. I will use science in many ways when I am an adult.
    c. Some concepts in science help me see how I relate to other people.
    d. When I leave school there will be many opportunities for me to use science.
    e. Science is very relevant to me.
3. **Index of self-efficacy in science**
    a. Explain why earthquakes occur more frequently in some areas than in others.
    b. Recognize the science question that underlies a newspaper report on a health issue.
    c. Interpret the scientific information provided on the labeling of food items.
    d. Predict how changes to an environment will affect the survival of certain species.
    e. Identify the science question associated with the disposal of garbage.
    f. Describe the role of antibiotics in the treatment of disease.
    g. Identify the better of two explanations for the formation of acid rain.
    h. Discuss how new evidence can lead you to change your understanding about the possibility of life on Mars.
4. **Index of self-concept in science**
    a. I can usually give good answers to test questions on school science topics.
    b. When I am being taught school science, I can understand the concepts very well.
    c. I learn school science topics quickly.
    d. I can easily understand new ideas in school science.
    e. Learning advanced school science topics would be easy for me.
    f. School science topics are easy for me
5. **Index of general interest in science**
    a. Human biology
    b. Topics in astronomy
    c. Topics in chemistry
    d. Topics in physics
    e. The biology of plants
    f. Ways scientists design experiments
    g. Topics in geology
    h. What is required for scientific explanations

6. **Index of enjoyment of science**
    a. I enjoy acquiring new knowledge in science.
    b. I generally have fun when I am learning science topics.
    c. I am interested in learning about science.
    d. I like reading about science.
    e. I am happy doing science problems.
7. **Index of instrumental motivation to learn science**
    a. I study school science because I know it is useful for me.
    b. Making an effort in my school science subject(s) is worth it because this will help me in the work I want to do later on.
    c. Studying my school science subject(s) is worthwhile for me because what I learn will improve my career prospects.
    d. I will learn many things in my school science subject(s) that will help me get a job.
    e. What I learn in my school science subject(s) is important for me because I need this for what I want to study later on.
8. **Index of future-oriented motivation to learn science**
    a. I would like to work in a career involving science.
    b. I would like to study science after secondary school.
    c. I would like to work on science projects as an adult.
    d. I would like to spend my life doing advanced science.
9. **Index of science-related activities**
    a. Watch TV programs about science
    b. Read science magazines or science articles in newspapers
    c. Visit web sites about science topics
    d. Borrow or buy books on science topics
    e. Listen to radio programs about advances in science
    f. Attend a science club
10. **Index of students' awareness of environmental issues**
    a. The consequences of clearing forests for other land use
    b. Acid rain
    c. The increase of greenhouse gases in the atmosphere
    d. Nuclear waste
    e. Use of genetically modified organisms (GMO)

Appendix C.

Statistical Results on Informal Learning Practices

Some national and international surveys have measured child and adult experiences in informal learning environments. The 2006 OECD PISA survey included a set of questions about the number of hours students spent in school and out of school studying language, mathematics and science (OECD 2011). That survey reports that US 15-year-olds spend about 1 hour a week in out of school learning of science, about the same as in other European countries (Table 1). About 13 percent of US 15-year-old students spent more than 2 hours a week studying in an after school science program. The OECD report finds that hours spent studying a subject of the formal school is related to higher PISA scores but not so in after school programs. The relationship of time spent studying and test scores in PISA is not supported by other evidence. For example, a recent report on science achievement in NAEP shows that students who had activities outside of school had higher scores. This result is also found in special analyses of the Longitudinal Survey of American Youth (LSAY) which indicate that students who visit museums and science centers have higher levels of science achievement during each of the school years. Special tabulations of the LSAY shows that the greatest amount of student activity in after school programs occurs during middle school and early secondary classes but declines between grade 7 and 10 from 40 percent to 20 percent reporting attending a museum or planetarium (similar trends during both summer and during school year). In adulthood, the survey respondents tend to report a significant level of attendance at informal learning centers often as parents of young children. Further analysis of these surveys could form the basis for a more serious detailed examination of the use of time in different learning settings and could inspire more thorough analyses of existing programs and assessments and recommendations for improvement in OST practices.

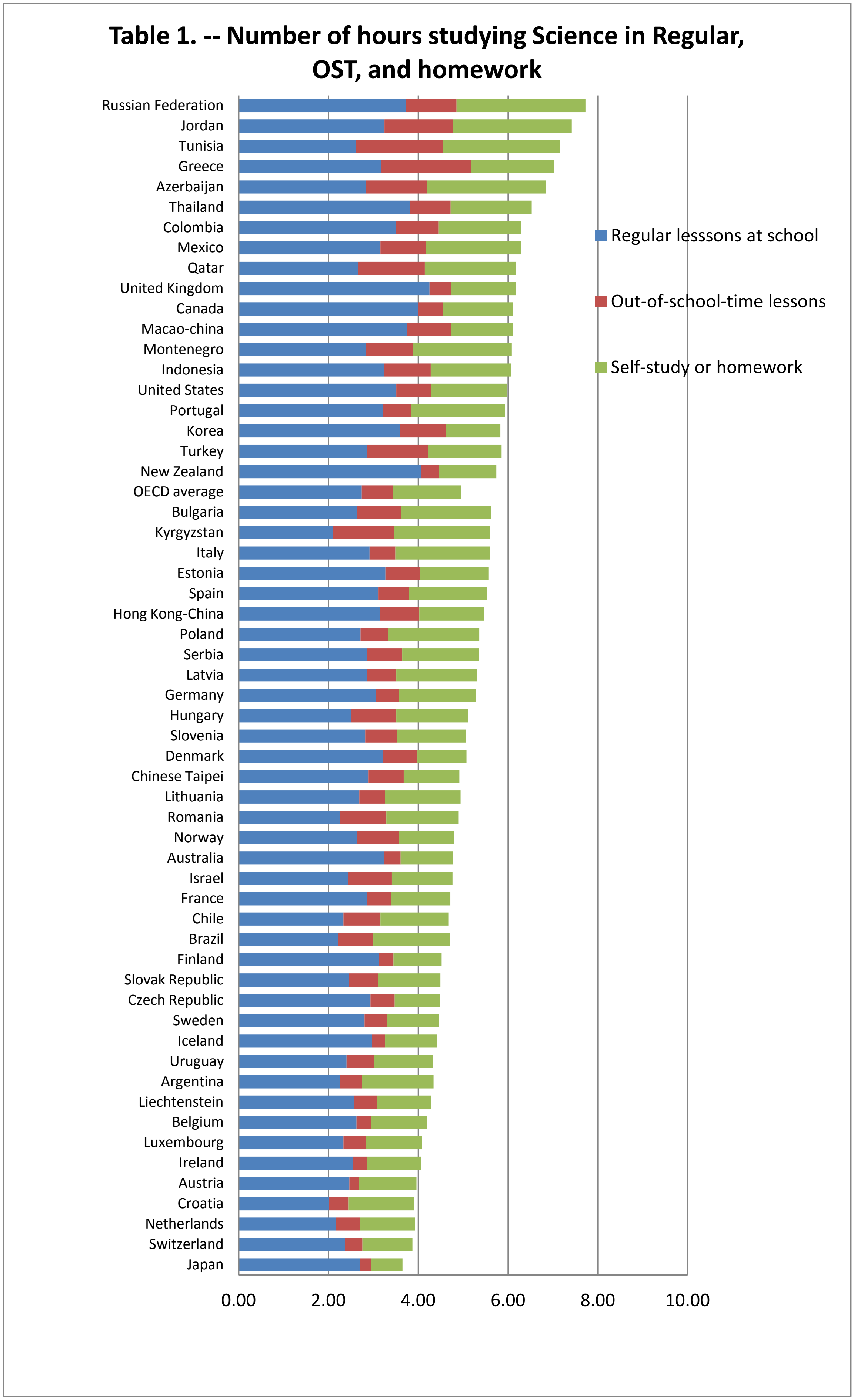
Table 1. -- Number of hours studying Science in Regular, OST, and homework
Russian Federation
Jordan
Tunisia
Greece
Azerbaijan
Thailand
Colombia
Mexico
Qatar
United Kingdom
Canada
Macao-china
Montenegro
Indonesia
United States
Portugal
Korea
Turkey
New Zealand
OECD average
Bulgaria
Kyrgyzstan
Italy
Estonia
Spain
Hong Kong-China
Poland
Serbia
Latvia
Germany
Hungary
Slovenia
Denmark
Chinese Taipei
Lithuania
Romania
Norway
Australia
Israel
France
Chile
Brazil
Finland
Slovak Republic
Czech Republic
Sweden
Iceland
Uruguay
Argentina
Liechtenstein
Belgium
Luxembourg
Ireland
Austria
Croatia
Netherlands
Switzerland
Japan
0.00
2.00
4.00
6.00
8.00
10.00
Regular lesssons at school
Out-of-school-time lessons
Self-study or homework

Source: Tabulations in: OECD (2011), *Quality Time for Students: Learning In and Out of School*, OECD Publishing. *http://dx.doi.org/10.1787/9789264087057-en*

Note:

This paper is based on personal experiences and reflections from a 30 year process of guiding the development and analyses of important national and international studies of student performances in many different conditions. I also spoke with Bill Schmidt, Ken Travers, John Dossey and Ina Mullis who were all involved with the Second and Third international IEA studies. My own experience involves engaging in studies of the definition of education, mathematics, science, reading, writing, and technology in both formal and informal settings. All of my experiences involved creating statistical measures and all of them were initiated with a discussion of the general principles of the subject matter.

REFERENCES

Ainley, M. (2006). Connecting with learning: Motivation, affect and cognition in interest

Bandura, A. (1986). Social Foundations of Thought and Action. Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1997). Self-efficacy: The Exercise of Control. New York: W.H. Freeman.

Baumert (Eds.), Interest and learning: Proceedings of the Seeon conference on interest and gender (pp. 91–104). Kiel, Germany: IPN .

Blumenfeld, P. C., Kempler, T. M., & Krajcik, J. S. (2006). Motivation and Cognitive Engagement in Learning Environments. In K. Sawyer (Ed.), The Cambridge handbook of the learning sciences (pp. 475-488). New York: Cambridge University Press.

Bracey, G. W. (1987). Why the schools can't improve: The upper limit hypothesis. *Journal of Instructional Development, 10,* 7□12.

Burstein, L. (1989). *Second International Study of Mathematics* (Vol. 3). London: Pergamon Press.

Coleman, J. S. (1975). Methods and results in the IEA studies of effects of school on learning. *Review of Educational Research, 45,* 335□386.

Coleman, J. S., & Associates. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Department of Health, Education, and Welfare.

Comber, L. C., & Keeves, J. P. (1973). *Science education in nineteen countries: An empirical study.* Stockholm: Wiley.

Corno, L. & Mandinach, E.B. (2004). “What we have learned about Student Engagement in the Past Twenty Years.” In D.M. McInerney & S. Van Etten (Eds.) Big Theories (pp. 297–326).

Deci, E. L. & Ryan, R. (2008b). Theoretical Overviews and Research Reviews. Retrieved from: http://www.psych.rochester.edu/SDT/publications_browse.php

Deci, E.L. & Ryan, R.M. (2008a). Self-determination theory: A macrotheory of human motivation, development, and health. Canadian Psychology, 49(3), 182-185.

Durik, A. M., & Harackiewicz, J. M. (2007). Different strokes for different folks: How personal interest moderates the effects of situational factors on task interest. Journal of Educational Psychology, 99, 597–610 .

Dweck, C.S. (2000). *Self Theories: Their Role in Motivation, Personality, and Development.* Philadelphia: Psychology Press.

Eccles, J. S., Adler, T. F., Futtermann, R., Goff, S. B., Kaczala, C. M. Meece, J. L., & Midgley, C. (1983). Expectancies, values and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives:Psychological and sociological approaches* (pp. 75–145). San Francisco CA: Freeman.

Eccles, J. S., Wigfield, A., Flanagan, C., Miller, C., Reuman, D., & Yee D. (1989). Self-concept, domain values, and self-esteem: Relations and changes at early adolescence. *Journal of Personality*, 57, 283–310 .

Eccles, J., Brown, B.V., & Templeton, J. (2008). A developmental framework for selecting indicators of well-being during the adolescent and young adult years. In B.V. Borwn (Ed.), *Key Indicators of Child and Youth Well-being* (pp. 197-236). New York, NY: Lawrence Erlbaum Associates.

Eliot, A.J. & Dweck, C.S. (2005) *Handbook of Competence and Motivation*. New York: Guilford Press.

Frenzel, A. C., Thrash, T. D., Pekrun, R., & Goetz, T. (2007). Achievement emotions in Germany and China: A cross-cultural validation of the Academic Emotions Questionnaire–Mathematics. *Journal of Cross Cultural Psychology*, 38, 302–309 .

Freudenthal, H. (1978). *Weeding and sowing: Preface to a science of mathematical education*. Amsterdam: Reidel.

Garden, R. (1987a). *The opportunity to learn.* Washington, DC: National Center for Education Statistics.

Garden, R. (1987b). Second IEA mathematics study. *Comparative Education Review, 31,* 47□68.

Gustafsson, J.-E., & Undheim, J. O. (1996). Individual differences in cognitive functions. In D. Berliner & R. Calfee (Eds.), *Handbook of research in educational psychology* (pp. 243□310). New York: Macmillan.

Harackiewicz, J.M. & Hulleman, C.S. (2010). The importance of interest: The role of achievement goals and task values in promoting the development of interest. *Social and Personality Psychology Compass*, 4: 42-52.

Harnqvist, K. (1987). The IEA revisited. *Comparative Education Review, 31,* 48□55.

Heckhausen, J. & Dweck, C.S. (1998) *Motivation and Self Regulation Across the Life Span*. Cambridge: Cambridge University Press.

Heyneman, S. P., & Loxley, W. A. (1982). Influences on academic achievement across high and low income countries: A re-analysis of IEA data. *Sociology of Education, 55,* 13□21.

Hidi, S. (1990). Interest and its contribution as a mental resource for learning. *Review of Educational Research*, 60, 549–571 .

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. Educational Psychologist, 41, 111–127 s15326985ep4102_4

Husen, T. (1979). An international research venture in retrospect: The IEA surveys. *Comparative Education Review, 23,* 371□385.

Husen, T. (1987). Policy impact of IEA research. *Comparative Education Review, 31,* 29□47.

Husen, T. (Ed.). (1967). *International study of achievement in mathematics* (Vol. 1). New York: Wiley.

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A . (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. Child Development 73, 509–527

Maltese, A.V. Tai, R.H. (2010). Eyeballs in the Fridge: Sources of early interest in science. *International Journal of Science Education. Vol. 32, No. 5, 15 March 2010, pp. 669–685.* ISSN 0950-0693 (print)/ISSN 1464-5289 (online)/10/050669–17.

McKnight, C., Crosswhite, F., Dossey, J., et al. (1987). *The underachieving curriculum: Assessing U.S. school mathematics from an international perspective*. Champaign, IL: Stipes.

Miller, J. D. & Brown, K. G. 1992a. The development of career expectations by American youth. In W. Meeus et. al. (Eds.), *Adolescence, careers, and cultures*. Berlin: Walter de Gruyter.

Miller, J.D. 1997. Civic scientific literacy in the United States: A developmental analysis from middle-school through adulthood, in Gräber, W. and Bolte, C. (Eds.), Scientific Literacy. Kiel, Germany: University of Kiel, Institute for Science Education. Pp. 121-142.

Miller, J.D. 1998. “The measurement of civic scientific literacy”. *Public Understanding of Science* 7:1-21.

Miller, J.D. 2000. The Development of Civic Scientific Literacy in the United States, in Kumar, D.D. & Chubin, D. (Eds.), Science, Technology, and Society: *A Sourcebook on Research and Practice*. New York: Plenum Press. Pp. 21-47.

Mosteller, F., & Moynihan, D. P. (1972). *On equality of educational opportunity.* New York: Random House.

Mullis, I., Martin, M., Beaton, A., Gonzalez, E., Kelly, D., & Smith, T. (1998). *Mathematics and science achievement in the final year of secondary school: IEA□s Third International Mathematics and Science Study (TIMSS*). Boston: Center for the Study of Testing, Evaluation, and Education Policy, Boston College.

National Assessment Governing Board (NAGB). (2007). ***Science Assessment and Item Specifications for the 2009 National Assessment of Educational Progress***. Pre-publication copy: Retrieved from http://www.nagb.org/publications/frameworks.htm.

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: U.S. Government Printing Office.

National Research Council. (2007). *Taking Science to School: Learning and Teaching Science in Grades K-8.* Committee on Science Learning, Kindergarten Through Eighth Grade. R.A. Duschl, H.A. Schweingruber, and A.W. Shouse (Eds.). Board on Science Education, Center for Education. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Research Council. (2009). *Learning Science in Informal Environments: People, Places, and Pursuits.* Committee on Learning Science in Informal Environments. P. Bell, B. Lewenstein, A.W. Shouse, and M.A. Feder (Eds.). Board on Science Education, Center for Education. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas.* Committee on a Conceptual Framework for New

K-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Science Board. 2010. *Science and Engineering Indicators 2010*. Arlington, VA: National Science Foundation (NSB 10-01).

National Science Board. 2012. *Science and Engineering Indicators 2012*. Arlington VA: National Science Foundation (NSB 12-01).

OECD. (2007) *PISA 2006 Science Competencies for Tomorrow's World*.**Volume 1 – Analysis.**

OECD (2010), *PISA 2009 Results: Learning to Learn – Student Engagement, Strategies and Practices (Volume III).* *http://dx.doi.org/10.1787/9789264083943-en*

OECD (2011), *Quality Time for Students: Learning In and Out of School*, OECD Publishing. *http://dx.doi.org/10.1787/9789264087057-en*.

Peaker, G. (1975). *An empirical study of education in twenty-one countries: A technical report*. New York: Wiley.

Pintrich, P.R. & Schunk, D. (2002). *Motivation in Education: Theory, Research, and Applications*. Merrill Prentice Hall. New Jersey.

Renninger, K. A. (2000). Individual interest and its implications for understanding intrinsic motivation. In B. Sansone & J. M. Harackiewitz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance* (pp. 373–404). New York, NY: Academic Press

Robitaille, D., Schmidt, W., Raizen, S., McKnight, C., Britton, E., and Nicol, C. (1993). *Curriculum Frameworks for Mathematics and Science.* Pacific Education Press.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55, 68-78.

Rychen, D.S., Salganik, L.H. (2001). Defining and Selecting Key Competencies. Hogrefe & Huber.

Schmidt, W. H., Jakwerth, P. M., & McKnight, C. C. (1998). Curriculum-sensitive assessment: Content does make a difference. *International Journal of Educational Research, 29*(6).

Schmidt, W. H., Jorde, D., Cogan, L., Barrier, E., Gonzalo, I., Moser, U., Shimizu, K., Valverde, G., McKnight, C., Prawat, R., Wiley, R., Raizen, S., Britton, T., & Wolfe, R. (1996). *Characterizing pedagogical flow: An investigation of mathematics and science teaching in six countries*. Dordrecht, the Netherlands: Kluwer.

Shavelson, R. J., Roeser, R. W., Kupermintz, H., Lau, S., Ayala, C., Haydel, A., Schultz, S., Quihuis, G., & Gallagher, L. (2002). Richard E. Snow's remaking of the concept of aptitude and multidimensional test validity: Introduction to the special issue. *Educational Assessment*, 8(2), 77–100.

Snow, R. E., Corno, L., & Jackson, D. N., III. (1996). Individual differences in affective and conative functions. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 243-308). New York: Simon & Schuster.

Sternberg, R. (2005). "Intelligence, Competence, and Expertise." *Handbook of Competence and Motivation*. Chapter 2.

Tai, R.H., Liu, C.Q., Maltese, A.V., & Fan, X. (2006). Planning early for careers in science. *Science*, 312: 1143-1144.

Travers, K. J., & Westbury, I. (1990). *The IEA study of mathematics I: Analysis of mathematics curricula*. Oxford, England: Pergamon Press.

Walker, D. A. (1976). *The IEA six subject survey: An empirical study of education in twenty-one countries.*

Wigfield, A., & Cambria, J. M. (2010a). Expectancy-value theory: Retrospective and prospective. In S. A. Karabenick & T. Urdan (Eds.), *Advances in motivation and achievement* (Vol. 16a) (pp. 35-70). Emerald Group.