Imagining Education Tailored to Assessment As, For, and Of Learning: Theory, Standards, and Quality Improvement

William P. Fisher, Jr.
University of California, Berkeley

Abstract

The demands of a new economic and technological age require leaders in education to awaken to new challenges, and learn to manage new responsibilities. Assessment will play a key role in meeting these challenges and in taking on these responsibilities. Assessment will not, however, be implemented in just the traditional sense of assessing learning for accountability purposes but will itself become a medium embodying and setting the stage for learning. Educators have made distinguished contributions in aligning learning progressions with assessments for use in self-directed individualized instruction. Highly technical principles and mathematical operations have been made accessible to end users who do not have the training or skills needed to directly employ those principles and operations themselves. Accordingly, the time has come to draw out some of the developments likely to follow from or impact the technical advances in assessment as, for, and of learning that are being put in practice. Three areas have the greatest promise for the future. They are theory development, standardized units of measurement, and systematic quality improvement methods. Putting these in place around self-directed learning-oriented assessment practices offers real hope for significant positive changes in education.

Introduction

We are in a new economic and technological age. Leaders in education who manage school systems and individual schools must awaken to new challenges, and to being accountable for new responsibilities. They must invent new educational outcome products, and, most importantly, they must match the supply of those products with growing demand in the world for responsible citizens with productive skills.

Teamwork and collaboration are increasingly required in industries that thrive on innovation. Schools need to teach more than just basic and critical thinking skills, they need to foster emotional intelligence and social skills, too. Schools must provide environments in which all of these skills are incorporated into day-to-day practices.

Passive reading and listening to lectures are far less effective for learning than self-directed participatory involvement and dialogue. Because they can support the latter with highly effective tools, online distance education and e-learning methods are being shown just as or more effective for learning as traditional classrooms.

Already today, the school a student attends matters less than what that student knows and can do. Leaders in education should be doing everything they can to make the market for educational outcomes as efficient as possible. Schools are partnering with employers, linking educational and training opportunities with recruitment efforts.

Assessment is playing a key role in meeting these challenges and in taking on these responsibilities. The changes taking place are paradigm-shifting (Gipps, 1994; Shepard, 2000). Assessment is not, however, being implemented in just the traditional sense of assessing learning for accountability purposes like grades, graduation, admissions, certification, or licensure. Instead, assessment has itself become a medium embodying and setting the stage for learning

(Black & Wiliam, 2009; Cheng & Mok, 2007; Earl, 2013; Feng, Heffernan, Koedinger, 2009; Ma, 2012; Mok, 2011; Wilson, 2009). Implementations integrating assessment and instruction in the classroom build on the large effect sizes observed in research on what works in education (Hattie & Timperley, 2007).

The research aligning learning progressions with assessments for use in self-directed individualized instruction is complex and technical. Putting the right methods and tools in the hands of the teachers and students – methods like self-directed learning strategies and tools like the S-P Chart, Item-Person Map, Kid-Map, and Unexpected Persons Map – will require innovative approaches but will have lasting effects (Mok, 2011).

The accomplishment is one in which highly technical principles and mathematical operations have been made accessible to end users who do not have the training or skills needed to directly employ those principles and operations themselves. The philosopher, Whitehead, observed that “Civilization advances by extending the number of important operations which we can perform without thinking about them.” Indeed, everyday tools like telephones, computers, and automobiles are now so complex that even engineering experts do not have the range of knowledge needed to master all of the component parts in a single device.

The question arises, then, as to what developments might follow from or further impact the technical advances in assessment as, for, and of learning that are being put in practice. There are three areas in which research and development seem to me to have the greatest promise for the future. The three substantive areas that will have a direct bearing on self-directed learning-oriented assessment are:

- Theory, which allows for greater efficiency in item development, test assembly, adaptive administration, instrument calibration, and measure estimation.

- Standardized units of measurement, which will facilitate universal comparability and maximize the involvement of all stakeholders (teachers, colleagues, students, parents, the community at large, employers, researchers, etc.) in the outcomes produced.
- Quality improvement methods that systematically implement the common metrics in a variety of environments, building up from the students' self-directed learning to self-directed improvement efforts at weekly faculty meetings, at quarterly parent-teacher meetings, at district-wide teacher conferences, etc.

In these remarks, advances in computer hardware, software and networks are taken for granted as the mechanisms by which we will have ever faster connections and processing speeds. What are not obvious and cannot be taken for granted, however, are the paths along which theory, standardized units of measurement, and quality improvement methods will develop. Though these paths can be sketched now, their full shape and direction remain to be determined. The best way to predict the future is to invent it, and so the best place to begin is where integrated assessment and instruction is now.

Assessment As and For Learning

Where does education begin? The most important realization about education is that we learn through what we already know. That's why early education focuses on language and numbers. Alphabets, characters, words, grammar, numbers, symbols, phonemes, etc. are the media of learning, so they, along with basic social and self-management skills, have to be learned before anything else.

To figure out what to learn next, students and teachers need to know what is already known. Assessment makes its point of entry in providing this information. It might do so in a manner as informal as a short conversation, or as structured as an adaptive online assessment. Advanced measurement modeling helps ensure that the information obtained is precise and meaningful, and advanced cognitive theory relates the assessment results to broader performance expectations.

Assessment as and for learning tells the individual student what to focus on next: the easiest lesson in the learning progression that has not yet been mastered. Here assessment results converge with the curriculum. Theoretical expectations as to developmental sequences align with practical experience in the way the difficulties of items and tasks increase with their cognitive complexity. This correspondence provides the structure by which formative feedback can individualize instruction and improve outcomes (Black & Wiliam, 1998, 2009; Hattie & Timperley, 2007; Wilson, 2004, 2009).

Assessment can be the place where learning happens, but the fact that students learn while answering test questions has often been viewed as more of a problem than an opportunity. Assessments deliberately designed to tap existing knowledge can leverage that knowledge to create new learning. Decades of research in measurement and cognition are coming to fruition in practical classroom applications (Feng, Heffernan, & Koedinger, 2009; Hannafin & Foshay, 2008; Law & Leung, 2012; Mok, 2011). As these results are brought to bear, demand will likely grow for better theoretical control, more efficient and meaningful communication, and systematic improvement methods. Leaders in education must awaken to new challenges beyond assessment, and learn new responsibilities.

Theory

“There is nothing so practical as a good theory” (Lewin, 1951, p. 169). Theory is efficient. Item difficulties are increasingly accurately predicted from previous experience or from a thorough understanding of the construct's properties (Embretson, 1984, 1998; Embretson & Daniel, 2008; Gierl & Lai, 2012; Stenner, Fisher, Stone, & Burdick, 2013; Stenner & Smith, 1982; Stenner & Stone, 2010). In this context, data gathering and analysis can become a needless waste of resources. Clocks, thermometers, voltmeters, and electrical cable are calibrated from theory, not data. The practical value of predictive theory is such that there would be no electrical industry or array of convenient consumer electronics and appliances if the resistance properties of every meter of cable had to be empirically calibrated.

As the patterns observed in assessment data repeat themselves over time and space and across millions of students and thousands of items, the difficulties of assessment items will be determined less by data analysis than by means of experimentally validated predictive theory. Table 1 lists a small sample of studies, their constructs and predictive successes. Figure 1 shows a typical scatterplot from one of those studies illustrating the correspondence between observed and expected item calibrations. Plainly test items automatically generated from theory (Bejar, Lawless, Morley, Wagner, Bennett, & Revuelta, 2003; Embretson, 1999; Gierl & Lai, 2012; Gorin & Embretson, 2012; Stenner, Swarz, Hanlon, & Emerson, 2012; Ying & Yang, 2008) stand to improve assessment efficiency by an order of magnitude.

Recent estimates put the cost of developing a single high stakes item for use in a pencil and paper assessment at US$4,000 (Stenner, et al., 2012). Items built by computers on the fly in the course of administering an assessment, in contrast, cost less than US$0.01. Maintaining test security is complex and difficult, introducing another array of costs. Compromised assessment items are useless when high stakes decisions are riding on the results. Single-use theory-based items make every assessment unique without compromising comparability, making response key cheating a thing of the past. Theory-informed, targeted curriculum materials (Law & Leung, 2012; Stenner, et al., 2012) can individualize instruction and assessment over time, and provide a basis for learning growth modeling (Figure 2). If educators dream big and take perfection as a goal, even the two sigma problem (Bloom, 1984) may not be out of reach (more on this below).

Standardized Units

Universally uniform units of measurement provide common languages used by communities of research and practice to coordinate their collective learning. Metrology, the science of calibrating instruments to standards, is increasingly credited with establishing the basis for distributed cognition across social networks, and so also for much of the power of science and commerce (Ashworth, 2004; Latour, 1987, pp. 147-157). In the same way that assessment tells students where they are and what to do next, so, too, it should allow teachers to see where they stand relative to their peers in their instructional effectiveness. Standardized units of measurement are essential to creating the needed common languages (Fisher, 2009, 2012a, 2012b).

Communication is simplified when numbers mean the same thing no matter which particular assessment items were used and no matter which particular student in which particular grade at which particular school is measured. Though this simplification is the fundamental requirement of Rasch's (1960; Andrich, 2010; Wright, 1977, 1999; Wright & Mok, 2000) concept of specific objectivity, the lack of predictive theory and the concomitant need for empirical instrument equating have made creating standard units of measurement impractical (Fisher & Stenner, 2013, p. 8). Theory-referenced standard units of measurement make it possible to plot longitudinal trajectories over time in growth charts. Growth in learning can be managed with the end in mind, so that action can be taken as soon as there is any sign of special need (Figure 2).

Consensus standards for uniform product definitions and units of measurement are of huge economic value. Measurement standards have been shown to increase productivity, reduce transaction costs, improve efficiency in research and development, enable the creation of new markets, and enhance product quality (NIST, 2009). For instance, according to the ISO (2010), the benefits of standards within the worldwide automobile industry contribute US$25 to $55 billion to the global economy annually.

Similar new efficiencies can likely be expected to accrue in education as measurement becomes more meaningful in terms of representing real change in universally uniform terms (Fisher, 2011). Educators will develop shared expectations, vocabularies, and terminologies around their own product definitions, such as the gains typically made in any given level of mathematics or reading instruction. These common languages will make it possible for educators to see more clearly what works and what does not, while also informing parents, students, and the public about the range of outcome quality available in local education markets. The ability to compare outcomes simply will be an important step forward in bringing the cost-quality relationship in education under better control.

Quality Improvement

The emerging culture of learning (Shepard, 2000) will increasingly blend with continuous quality improvement values and methods (Deming, 1986, 1994; Heinemann, Fisher, & Gershon, 2006; Lunenberg, 2010) and the ideal of learning organizations (Senge, 2006; Senge, Kleiner, Roberts, Ross, & Smith, 1994; Kotter, 1996). Strong theory, standardized outcome definitions, and uniform units of measurement set the stage in education for powerful implementations of the ideas that transformed the manufacturing and service industries. Deming's 14 Total Quality Management (TQM) principles and four points on quality have a wide scope of application relevant in every respect to the needs for improved outcomes in education (see boxes).

In accord with TQM and CQI methods, it will be essential to separate general and systemic common causes of quality issues managed by school leadership from local and idiosyncratic special causes of quality issues managed by students and teachers. Traditional quality control methods focus on removing undesired results from production without addressing the systemic causes bringing them about in the first place. These so-called "tail-chopping" methods get their name from the process of cutting off the low-quality end of a distribution by focusing on worker training or local, special causes of problematic results. The problem with this approach is that the systemic causes of poor quality remain unaddressed, and nothing is done to prevent the tail of the quality distribution from simply reappearing in the next round of production.

TQM and Continuous Quality Improvement (CQI) methods, in contrast, focus on shifting the entire curve to a higher overall level. This is accomplished by modifying the general, systemic causes of problems to improve the process as a whole.

In education, Bloom (1984) grasped the essence of TQM and CQI curve-shifting methods in his definition of what he called the two sigma problem. To improve group instruction to the point that its outcomes are equivalent to the outcomes of one-on-one tutoring, it is necessary to move the average performance up the scale by two standard deviations. Figure 3 shows the distributions commonly associated with classroom instruction, mastery learning, and tutoring. Bloom focused so intently on the two sigma problem that Ben Wright, speaking at Bloom's retirement celebration, referred to these curves as Bloom's "personal logo" (Bloom, 2006, p. xvi). The convergence within education of networked information technology, strong theory, standardized metrics and outcome definitions, and systematically applied TQM/CQI methods may in time fulfill Bloom's dream of achieving the best possible outcomes for all students.

Moving into the Future

As the research and technical foundations of assessments for and as learning are laid, new and broader issues of academic achievement will arise. Assessment for and as learning structures academic planning and goal setting, but both of these have been shown to have less of an effect on achievement than students' academic goal orientation (Zhu & Mok, 2012). Alignment with mastery goals and studying for the love of learning require students to fully possess their reasons and purpose in pursuing knowledge.

And what goes for students goes double for teachers and educational leadership. Fulfilling every student's potential will demand that educators take ownership of the full scope of the educational process, from beginning to end. Figure 4 illustrates the operational change and improvement cycle proven in its effectiveness time and again across a wide range of industries. Note the correspondence between predictive theory and guiding ideas, between standards and infrastructural innovations, and between TQM/CQI and applied theory, methods, and tools. Organizations capable of creating a culture of continuous learning around Deming's 14 principles will endlessly move through the change and improvement cycle to create and share new value.

Figure 5 shows the conditions that must be met for sustainable change to be realized, along with the negative results that accrue when any given element is omitted (Knoster, Villa, & Thousand, 2000). Vision, skills, incentives, resources, and a plan must all be in place to avoid confusion, anxiety, resistance, frustration, and treadmills. Creating a culture of learning with an advanced information infrastructure, predictive theory, the common languages of standard metrics, and systematic quality improvement methods is a huge challenge, but as Georg Rasch (1980, p. xx) recognized in a related context, "once the problem has been formulated it does seem possible to meet it."

Table 1 : Publications Featuring Item Calibrations Predicted from Theory

Authors	Scale	Variance Explained
Irvine, Dunn, Anderson, 1990	British Army Recruitment Battery	R > .70
Embretson, 1998	Abstract Reasoning Test	R > .70
Stenner & Smith, 1982	Knox Cube Test	R > .90
Fischer, 1973	Elementary calculus test	R > .85
Stenner, et al, 1983	Peabody Vocab Test	R > .80
Stenner, et al, 1997	Reading tests	R > .90
Bejar, et al, 2003	Math tests	R > .85
Fisher, 2008	Physical function surveys	R > .90

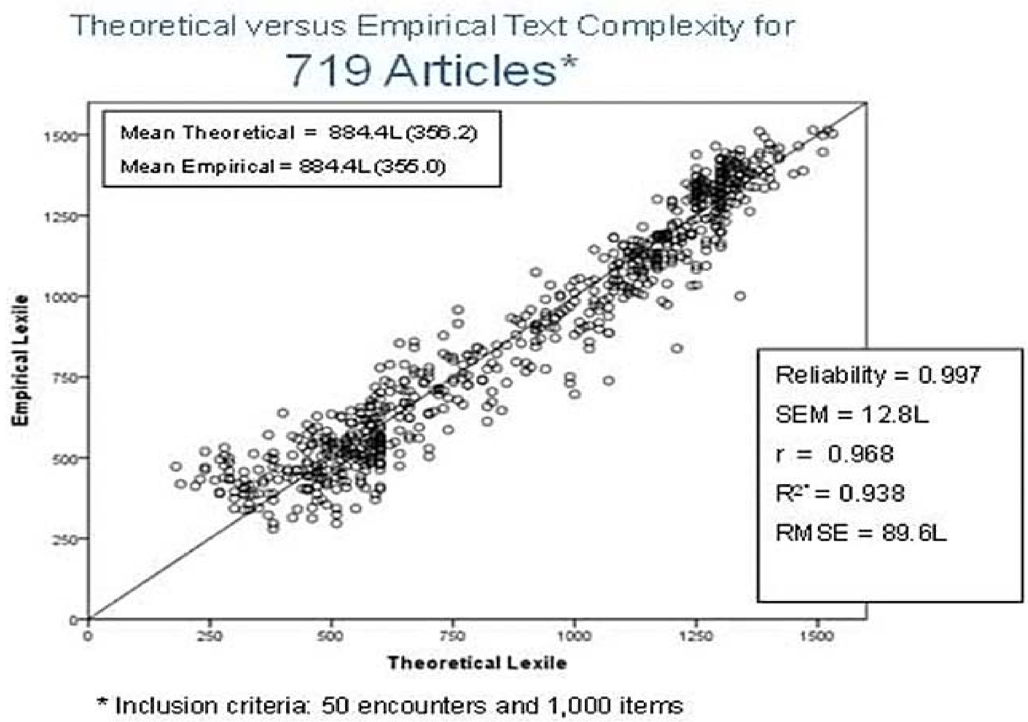

Figure 1. Empirical vs. theoretical item calibrations
(Stenner & Burdick, 1997)

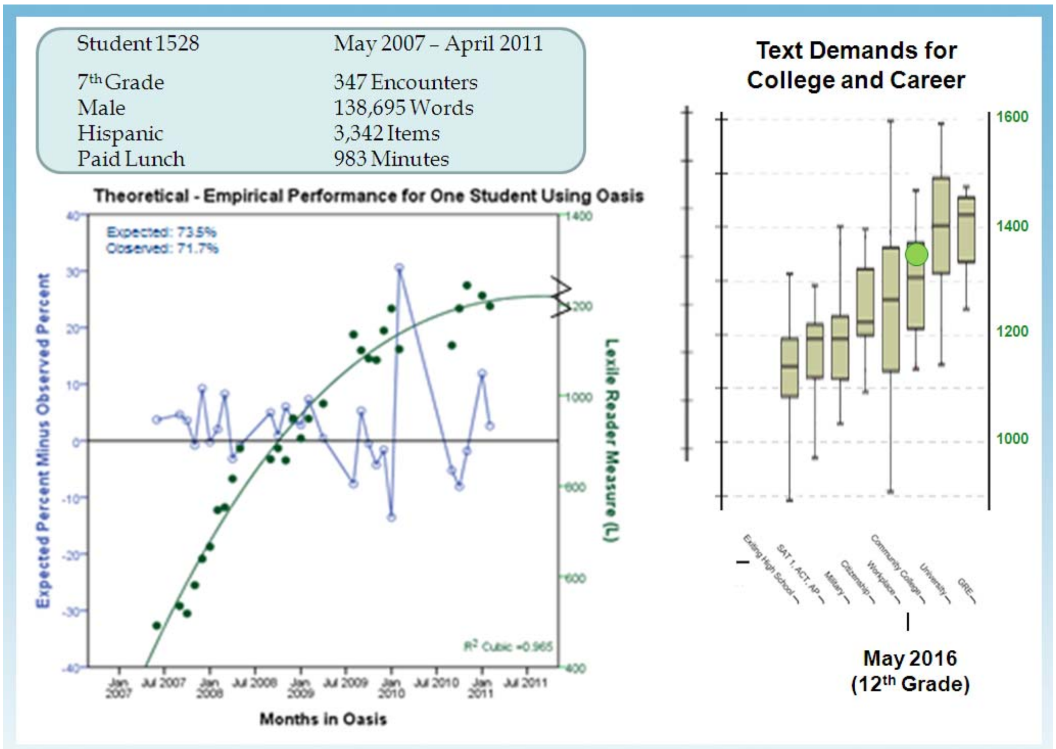

Figure 2. Individualized growth in reading relative to desired outcomes (Stenner, Swartz, Hanlon, & Emerson, 2012)

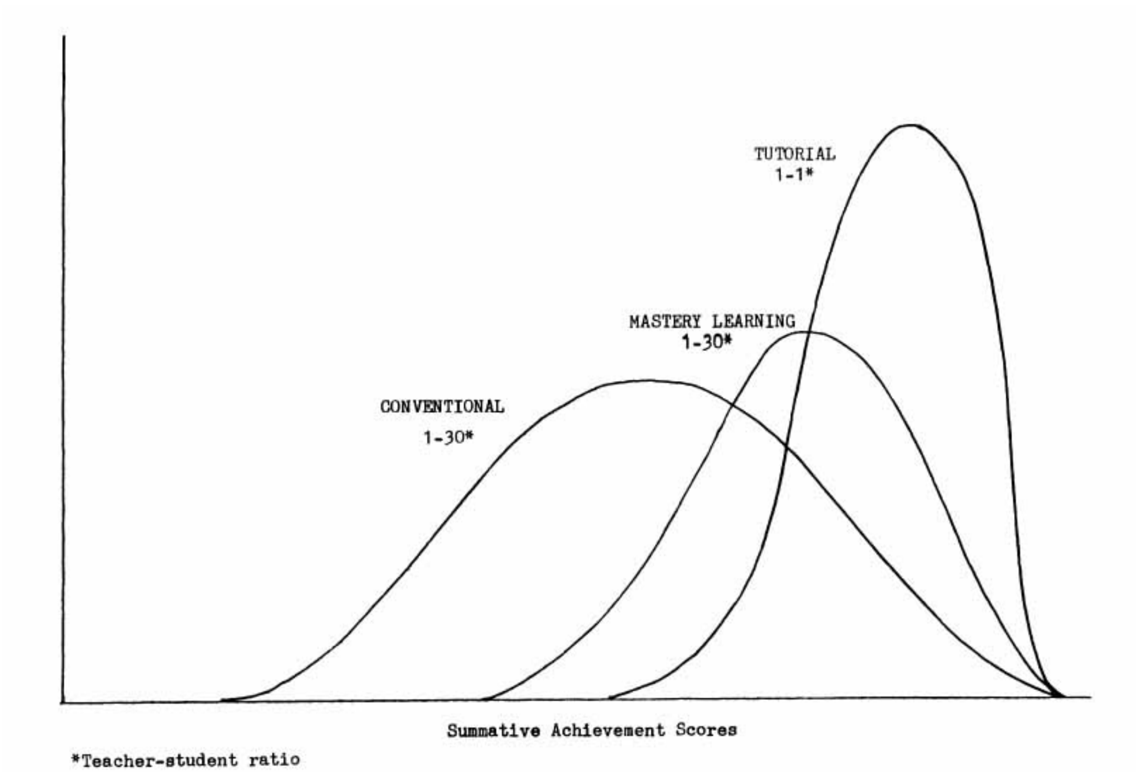

Figure 3. The two sigma problem (Bloom, 1984, p. 5)

Deming's 14 Total Quality Management Principles Adapted to Education
(Lunenberg, 2010; Deming, 1986, 1994)

1. Constancy of purpose
2. Adopt the new philosophy.
3. Augment end-point accountability with formative assessment.
4. Focus less on lowest-cost bids and more on long term relationships of trust.
5. Improve constantly and forever.
6. Institute continuous on-the-job training.
7. Lead toward learning for all.
8. Drive out fear.
9. Break down barriers between staff areas.
10. Eliminate slogans, exhortations, and productivity targets, like proficiency levels.
11. Eliminate numerical quotas and goals.
12. Remove barriers that rob students and teachers of their natural pride in their work.
13. Institute vigorous education and training for all.
14. Involve everyone in the transformation.

Deming’s 4 Points Brought to Bear in Education
(Adapted from Lunenberg, 2010; Deming, 1986, 1994)

1. Appreciate the system: Understand all of the processes and roles relevant to how educational outcomes are produced, including teaching methods, relationships with textbook and material suppliers and producers, the students and their families, the community, teachers, support staff, and the material consequences of the building housing the school, the cafeteria operations, and the physical plant.

2. Knowledge of variation: Understand the range and common vs. special causes of variation in quality, and use measures that can be interpreted and applied not just in accountability applications, but in assessment as and for learning.

3. Theory of knowledge: Understand what knowledge is and the limits of what can be known.

4. Psychology: Understand human nature so as to be able to create satisfying, joyful work for all involved.

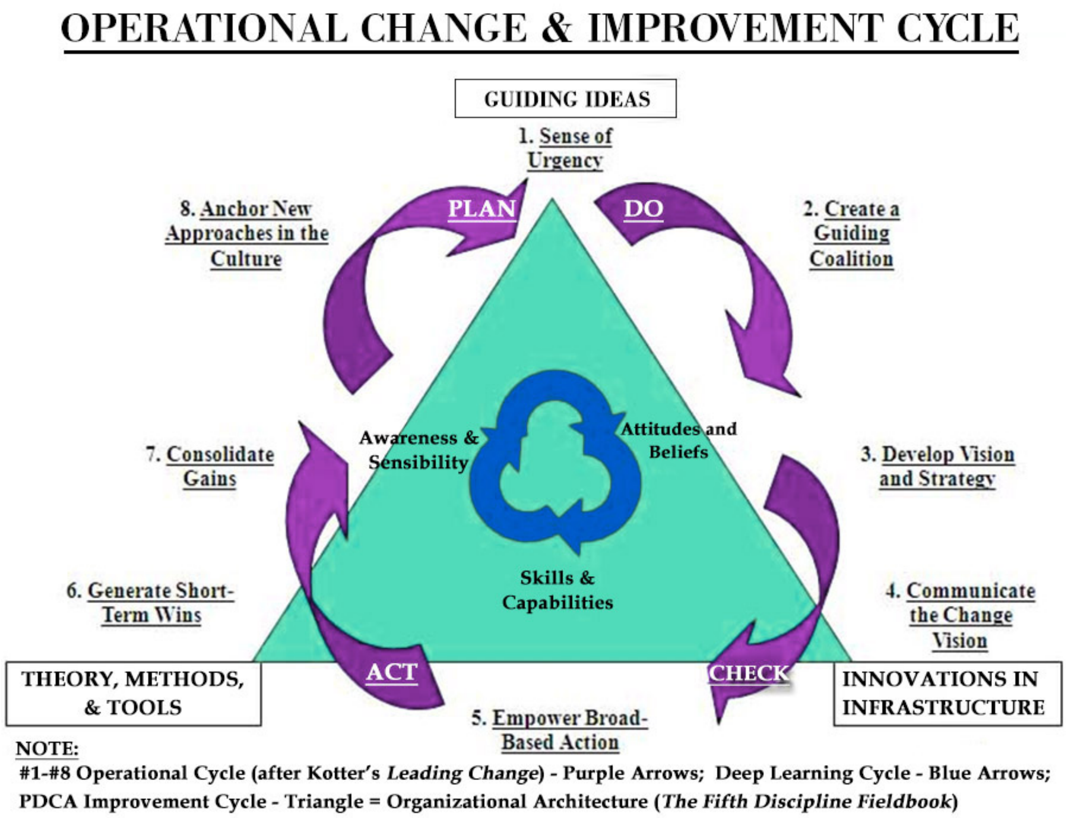

Figure 4. Operational change and improvement cycle (Kotter, 1996; Senge, 2006; Senge, Kleiner, Roberts, Ross, & Smith, 1994)

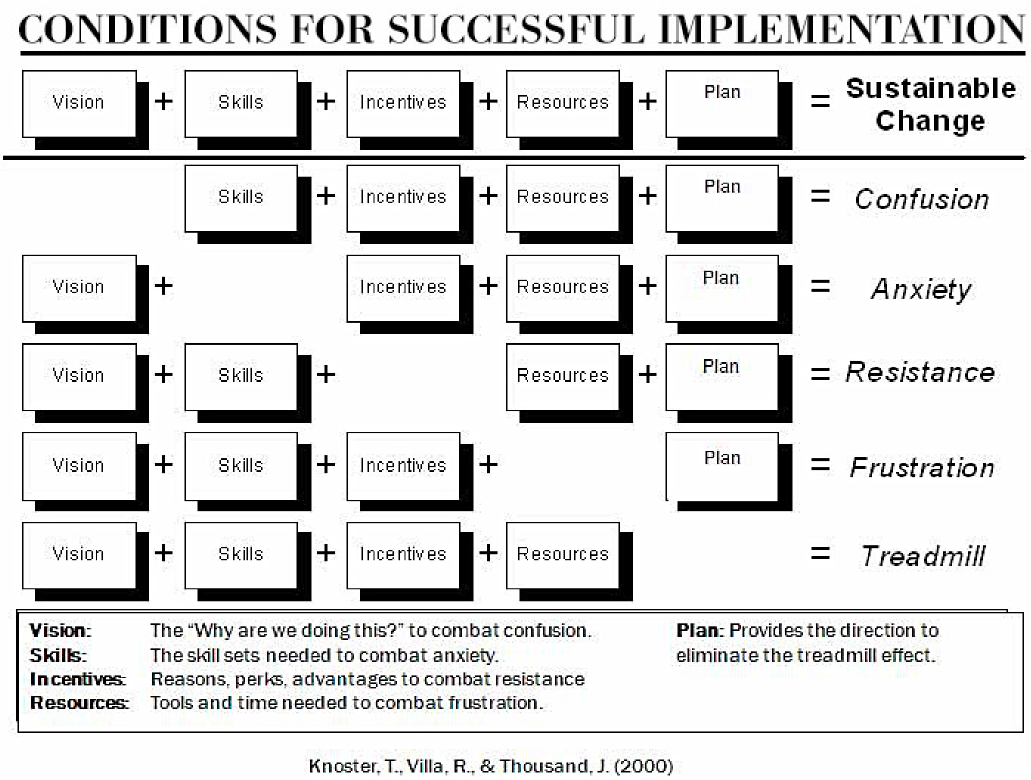

Figure 5. Conditions for successful implementation (Knoster, Villa, & Thousand, 2000)

References

Andrich, D. (2010). Sufficiency and conditional estimation of person parameters in the polytomous Rasch model. *Psychometrika*, *75*(2), 292-308.

Ashworth, W. J. (2004). Metrology and the state: Science, revenue, and commerce. *Science, 306*(5700), 1314-1317.

Bejar, I., Lawless, R. R., Morley, M. E., Wagner, M. E., Bennett, R. E., & Revuelta, J. (2003). A feasibility study of on-the-fly item generation in adaptive testing. *The Journal of Technology, Learning, and Assessment, 2*(3), 1-29.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-74.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*, 5-31.

Bloom, B. S. (1984). The two sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*(6), 4-16.

Bloom, S. (2006). Preface. In T. R. Guskey (Ed.), Benjamin S. Bloom: *Portraits of an educator* (pp. xi-xx). Lanham, Maryland: Rowman & Littlefield Education.

Cheng, Y. C., & Mok, M. M. C. (2007). School-based management and paradigm shift in education: An empirical study. *International Journal of Educational Management, 21*(6), 517-542.

Deming, W. E. (1986). *Out of the crisis*. Cambridge, MA: MIT Press.

Deming, W. E. (1994). *The new economics for industry, government, education* (2d Ed.). Cambridge, Massachusetts: MIT Center for Advanced Educational Services.

Earl, L. M. (2013). *Assessment as learning: Using classroom assessment to maximize classroom learning*. Thousand Oaks, California: Corwin.

Embretson, S. E. (1984). A general latent trait model for response processes. *Psychometrika, 49*(2), 175-186.

Embretson, S. E. (1998). A cognitive design system approach to generating valid tests: Application to abstract reasoning. *Psychological Methods*, 3(3), 380-396.

Embretson, S. E. (1999). Generating items during testing: Psychometric issues and models. *Psychometrika, 64*(4), 407-433.

Embretson, S. E., & Daniel, R. C. (2008). Understanding and quantifying cognitive complexity level in mathematical problem solving items. P*sychology Science Quarterly, 50*, 328-344.

Feng, M., Heffernan, N., & Koedinger, K. (2009). Addressing the assessment challenge with an online system that tutors as it assesses. *User Modeling and User-Adapted Interaction, 19*(3), 243-266.

Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica*, 37, 359-374.

Fisher, W. P., Jr. (2008). *A predictive theory for the calibration of physical functioning patient survey items*. Presented at the Second Conference on Patient Reported Outcomes Measurement Information Systems, Bethesda, Maryland: NIH and NIAMS.

Fisher, W. P., Jr. (2009). Invariance and traceability for measures of human, social, and natural capital: Theory and application. *Measurement, 42*(9), 1278-1287.

Fisher, W. P., Jr. (2011). Bringing human, social, and natural capital to life: Practical consequences and opportunities. *Journal of Applied Measurement*, *12*(1), 49-66.

Fisher, W. P., Jr. (2012a). Measure and manage: Intangible assets metric standards for sustainability. In J. Marques, S. Dhiman & S. Holt (Eds.), *Business administration education: Changes in management and leadership strategies* (pp. 43-63). New York: Palgrave Macmillan.

Fisher, W. P., Jr. (2012b). What the world needs now: A bold plan for new standards. *Standards Engineering*, *64*(3), 1 & 3-5.

Fisher, W. P., Jr., & Stenner, A. J. (2013). On the potential for improved measurement in the human and social sciences. In Q. Zhang (Ed.), *Pacific Rim Objectiive Measurement Symposium 2012 Conference Proceeding* (pp. 1-11). Berlin, Germany: Springer-Verlag.

Gierl, M. J., & Lai, H. (2012). The role of item models in automatic item generation. *International Journal of Testing, 12*(3), 273-298.

Gipps, C. (1994). *Beyond testing*. Abingdon: Routledge.

Gorin, J. S., & Embretson, S. E. (2012). Using cognitive psychology to generate items and predict item characteristics. In M. J. Gierl & T. M. Haladyna (Eds.), *Automatic item generation: Theory and practice* (pp. 136-156). New York: Routledge.

Hannafin, R. D., & Foshay, W. R. (2008). Computer-based instruction's (CBI) rediscovered role in K-12: An evaluation case study of one high school's use of CBI to improve pass rates on high-stakes tests. *Educational Technology Research & Development, 56*(2), 147-160.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112.

Heinemann, A. W., Fisher, W. P., Jr., & Gershon, R. (2006). Improving health care quality with outcomes management. *Journal of Prosthetics and Orthotics, 18*(1), 46-50.

International Organization for Standardization (ISO). (2010). *Assessing economic benefits of consensus-based standards*. Retrieved 6 July 2013, from http://www.standardsinfo.net/info/benefits/Economic%20benefits%20of%20standardization-Overview.pdf.

Irvine, S. H., Dunn, P. L., & Anderson, J. D. (1990). Towards a theory of algorithm-determined cognitive test construction. *British Journal of Psychology, 81*, 173-195.

Knoster, T. P., Villa, R. A., & Thousand, J. S. (2000). A framework for thinking about systems change. In R. A. Villa & J. S. Thousand (Eds.), Restructuring for caring and effective education: *Piecing the puzzle together, 2nd Ed*. (pp. 93-128). Baltimore: Paul H. Brookes.

Kotter, J. (1996). *Leading change*. Cambridge, Massachusetts: Harvard Business School Press.

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. New York: Cambridge University Press.

Law, K. K. Y., & Leung, F. F. Y. (2012). Design of web-based support materials to address learning difficulties – vocabulary building skills. *Assessment and Learning, 1*, 165-182.

Lewin, K. (1951). *Field theory in social science: Selected theoretical papers* (D. Cartwright, Ed.). New York: Harper & Row.

Lunenberg, F. C. (2010). Total Quality Management applied to schools. *Schooling, 1*(1), 1-6.

Ma, M. K. K. (2012). Assessment for learning: Collaborative research and professional development projects in school context. *Assessment and Learning, 1*, 157-164.

Mok, M. M. C. (2011). The assessment for, of and as learning in mathematics: The application of SLOA. In B. Kaur & K. Y. Wong (Eds.), *Assessment in the mathematics classroom* (pp. 187-215). Singapore: World Scientific.

NIST. (2009). *Outputs and outcomes of NIST laboratory research*. Retrieved 1 March 2011, from NIST: http://www.nist.gov/director/planning/studies.cfm.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests* (Reprint, with Foreword and Afterword by B. D. Wright, Chicago: University of Chicago Press, 1980). Copenhagen, Denmark: Danmarks Paedogogiske Institut.

Senge, P. (2006). *The fifth discipline: The art & practice of the learning organization* (2nd Ed.). New York: Random House.

Senge, P., Kleiner, A., Roberts, C., Ross, R. B., & Smith, B. J. (1994). *The fifth discipline fieldbook: Strategies and tools for building a learning organization*. New York: Crown Business.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4-14.

Stenner, A. J., & Burdick, D. S. (1997). The objective measurement of reading comprehension. [Response to technical questions raised by the California Department of Education Technical Study Group], Durham, North Carolina: MetaMetrics, Inc.

Stenner, A. J., Fisher, W. P., Jr., Stone, M., & Burdick, D. (2013). Causal Rasch models. *Frontiers in Psychology*, in press.

Stenner, A. J., & Smith III, M. (1982). Testing construct theories. *Perceptual and Motor Skills, 55*, 415-426.

Stenner, A. J., Smith, M., III, & Burdick, D. S. (1983). Toward a theory of construct definition. *Journal of Educational Measurement, 20*(4), 305-316.

Stenner, A. J., & Stone, M. (2010). Generally objective measurement of human temperature and reading ability: Some corollaries. *Journal of Applied Measurement, 11*(3), 244-252.

Stenner, A. J., Swartz, C., Hanlon, S., & Emerson, C. (2012). *Personalized learning platforms*. Presented at the Pearson Global Research Conference, Fremantle, Western Australia.

Wilson, M. (Ed.). (2004). *Towards coherence between classroom assessment and accountability*. Chicago, Illinois: University of Chicago Press.

Wilson, M. (2009). Measuring progressions: Assessment structures underlying a learning progression. *Journal of Research in Science Teaching, 46*, 716–730.

Wright, B. D. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement, 14*(2), 97-116.

Wright, B. D. (1999). Fundamental measurement for psychology. In S. E. Embretson & S. L. Hershberger (Eds.), *The new rules of measurement: What every educator and psychologist should know* (pp. 65-104). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Wright, B. D., & Mok, M. M. C. (2000). Understanding Rasch measurement: Rasch models overview. *Journal of Applied Measurement, 1*(1), 83-106.

Ying, M.-H., & Yang, H.-L. (2008). Computer-aided generation of item banks based on ontology and Bloom's taxonomy. In F. Li, J. Zhao, T. Shih, R. Lau, Q. Li & D. McLeod (Eds.), *Advances in Web Based Learning - ICWL 2008* (Lecture Notes in Computer Science) (pp. 157-166). Berlin: Springer.

Zhu, J., & Mok, M. M. C. (2012). Effect of academic goal orientation, goal setting, and planning on academic achievement of secondary students in Hong Kong. *Assessment and Learning, 1*, 11-31.

Author's e-mail:　wfisher@berkeley.edu