Don’t Putter

There is every temptation for the teacher to putter. There is less temptation for the principal and superintendent, but whoever deals with children in an authoritative or autocratic way is tempted to putter.

Puttering in school is in the same class of activities as the aimless whittling of a loafer in a country store. There is real pleasure in whittling for the sake of whittling. It takes no skill to whittle as it does to make something. Whittling is a brainless activity, but it satisfies the desire to whittle.

--Journal of Education, New England and National, April 17, 1919
# INSIDE THIS ISSUE

**Editor's Corner**  
Peter Hendrickson, Ph.D.  
1

**Focus Issue: Evidence Based Practice in Education**  
Peter Hendrickson, Ph.D.  
2

**Five Hierarchies/Continuums/Types of Evidence for Evidence Based Practice**  
Robert Rayborn, Ph.D.  
3

**Commentary: A Medical Model of Educational Research Redux**  
Robert Rayborn, Ph.D.  
4

**Institute of Education Sciences Levels of Evidence for Practice Guides**  
Jessica Cohen, Ph.D., Jerry Johnson, Ph.D., and Kimberly Markworth, Ph.D.  
5

**Mathematics and Evidence-Based Practice: Using Clinical Research in a New Suit-of-Clothes**  
Brian Rick  
6

**Making it Simple: Four Essentials to Measuring Impact**  
Peter Hendrickson, Ph.D.  
7

**Evidence Based Focus: Resources for Special Education Programs**  
Donnita Hawkins  
8

**Evidence Based Instruction: Tools & Processes for Observing & Advancing Effective Instructional Teaching Practices**  
Jeannette LaFors, Ph.D. and Candace Gratama, Ed.D.  
9

**A Leadership Perspective on Implementing a STEM Initiative**  
Margery Ginsberg, Ph.D., Chris Kinsey and Julia M. Zigarelli  
10

**School Conditional Growth Model: How to Make the “Apples to Apples” Comparison Possible?**  
Yun Xiang, Ph.D. and Carl Hauser, Ph.D.  
11

**School Year 2009-10 Relationships Between Math Benchmark and State Test Scores**  
Peter Hendrickson, Ph.D. and Shannon Edwards  
12

**Data Visualizations with Google Gadgets**  
Bruce J. Denton  
13

**Techniques for Effective Use of Color in Data Display**  
Andrea Meld, Ph.D.  
14

**Dr. D’ Lema: A WERA Journal Forum on Data and Research Ethics**  
Andrea Meld, Ph.D.  
15

**Data Privacy: What it is and Why it Matters**  
Andrea Meld, Ph.D.  
16

**Book Review: Operational Best Practices for Statewide Large-Scale Assessment Programs**  
Frosyne Mensendick and J. Scott Taylor  
17

**Book Review: Beautiful Evidence by Edward Truete**  
Jack Monpas-Huber  
18

**Book Review: Brain Rules: 12 Principles for Surviving and Thriving at Work, Home and School by John Medina**  
Heather Rader  
19

**My Favorite Kosher Vegetarian Punjabi Restaurant**  
Andrea Meld, Ph.D.  
20

**Reviewer Acknowledgement Volume 3**  
21

## Editor’s Corner

### First Steps Towards A Peer Reviewed WERA Journal

This issue marks the transition of The WERA Educational Journal from editor reviewed and revised articles to (mostly) peer reviewed and revised papers. While most papers are invited by the editors, any author in the greater WERA community may submit papers for consideration. We've lengthened the time between manuscript submission deadlines and e-publishing dates to provide time for thoughtful, independent reviews and subsequent revisions prior to final acceptance of the manuscript.

For some reviewers this is new territory. I thank them for answering the call and providing prompt critiques to authors. For most of us, publication in peer reviewed journals is not key to advancement or recognition in our careers. We're deeply engaged with young learners and darn busy. This issue marks the first move to providing a choice of selecting the entire journal to read or simply a single paper or two. Photo editor Don Schmitz has designed and populated a new “front end” for the journal to provide a menu of offerings.

We hope that the journal can help bridge the gap between higher education and the classroom through the authorship and peer review process.

We're also aware that instruction is at the center of most of our work, not research or program evaluation. Partnerships with colleges, universities and other research organizations are not the norm. We're deeply engaged with young learners and damn busy. This issue marks the first move to providing a choice of selecting the entire journal to read or simply a single paper or two. Photo editor Don Schmitz has designed and populated a new “front end” for the journal to provide a menu of offerings.

Our next issue in December will focus on College and Career Readiness. Copy deadline for papers is September 23, 2011. If you wish to become a reviewer, please submit a letter of interest with attached CV. Guidance to reviewers is included in this issue. Letters to the editor are welcomed, as well.

---Peter Hendrickson, Ph.D.---
Focus Issue: Evidence Based Practice in Education

This issue of the Journal focuses on Evidence Based Practice (EPB) in education, in particular, in the Northwest. A medical model for evaluation evidence and guiding practice has developed since the early 1970's (Sackett et al, 1996). The reading wars led to a call for EPB in education, notably a requirement for randomized controlled trails (RCT) (USDOE, 2003).

Teachers, principals and others central to student learning are inherently constructivists--they create knowledge and understanding based on their own experiences. Left to our own devices, we naturally and legitimately accept and follow what scientists (education or other) classify as a lower level of evidence. In medicine that's expert opinion without explicit critical appraisal (EBOC, 2002) and in education we generally find evidence based on expert opinion not supported by strong (replicated RCTs) or moderate (non-replicated RCTs or quasi experimental research). Among the many hierarchies for classifying levels of research is Dalton's (2010) review of five with relevancy to special education.

Former Education Northwest Center for Research Director Bob Rayborn reflects on his call in WERA's The Standard Deviation newsletter nearly 20 years ago to adopt the medical model in education. His observation: Education simply does not have the same empirical footing as found in medicine.

Three Western Washington University math professors with close ties to teachers, Jessica Cohen, Jerry Johnson and Kimberly Markworth, look at EBP in mathematics education. They found scant evidence of the practice but offer two examples where EPB might flourish.

Donnita Hawkins, an OSPI special education coordinator, calls for special educators to use information about effective practices when selecting interventions. She identifies three sources for reviews and provides a side-by-side comparison of the PALS intervention findings.

Program evaluators Jeannette LaFors (Teachscape) and Candace Gratama (BERC Group) reflect on the use of Classroom Walk Throughs (CWTs) to gather evidence about effective instructional practices. They offer a case study of a predominantly Latino Washington district where CWTs prevail and are associated with improved achievement.

Margery Ginsberg (UW), Chris Kinney (CHS) and Julia Zigarelli (UW) offer a second case study of a STEM intervention at Cleveland High School in Seattle using qualitative methods to gather evidence.

Good doctors use both individual clinical expertise and the best available external evidence, and neither alone is enough. Without clinical expertise, practice risks becoming tyrannized by evidence, because even excellent external evidence may be inapplicable to or inappropriate for an individual patient. Without current best evidence, practice risks rapidly become out of date, to the detriment of patients. (p. 2)

References


--Editor
## Five Hierarchies/Continuums/Types of Evidence for Evidence Based Practice

<table>
<thead>
<tr>
<th>Author</th>
<th>Hierarchy</th>
</tr>
</thead>
</table>
Level I: Meta Analyses/Systematic Reviews  
Level II: Randomized Controlled Trial/Experimental/Quasi-Experimental  
Level III: Descriptive/Case Studies/Series  
Level IV: Expert Opinion  
Level V: Animal Research |
I. Experimental: Identical groups, randomly assigned to treatment, and control groups  
II. Quasi-Experimental: Treatment and control groups not randomly assigned, but appearing identical  
III. Correlational with statistical controls: Treatment and comparison groups not identical but statistics control for important differences  
IV. Correlational without statistical controls: Treatment and comparison groups different, but differences assumed not important. For use with large sample.  
V. Case studies: Only treatment group, and assumes differences among participants not important. For use with small sample. |
Level 1) Randomized experimental design or well-designed randomized control studies  
Level 2) Controlled studies without randomization (quasi-experimental designs)  
Level 3) Well-designed non-experimental studies (correlational and case studies)  
Level 4) Expert opinions (committee reports, consensus conferences, clinical experience of respected authorities) |
1) Randomized controlled trials  
2) Experimental & quasi-experimental studies  
3) Survey & correlational research (simple & multiple correlation, regression analysis, analysis of variance)  
4) Expert opinion (defining of processes, meanings, categories & practices by field professionals)  
5) Ethnographies/case studies/observations (analysis of consequences of activities by interaction/conversation/discourse)  
6) Ethics studies (universal vs. selective action, informed choices, social inequities, social justice, resource allocation, and values) |
a) experimental group  
b) correlational  
c) single subject  
d) qualitative designs |

---

Commentary: A Medical Model of Educational Research Redux
By Robert Rayborn, Ph.D.

“The times they are still a changing”
With all due respect to Bob Dylan

I recently revisited the contents of an article I authored in The Standard Deviation, almost twenty years ago entitled, A Medical Model of Educational Research (Rayborn, 1992). In that article, I advocated for the adoption of a medical model of empiricism for educators. Waxing prophetic at the time, I advanced the notion that, in the not too distant future, educators like doctors would have scientific evidence of effectiveness for the various programs they were considering. Practicing educators would be professionally obligated to select programs with the greatest documented positive effect. Failure to do so would expose them to lawsuits similar to doctors facing charges of malpractice. This then was the essence of the medical model concept.

That was then, and much of the article was speculation. This is now, and it might be good to take stock of the current state of affairs to see how prescient the article actually was after 20 years. During the last decade the focus of the U.S. Department of Education has clearly shifted in the direction of empiricism. For those of us working directly in this system, the term strident empiricism would not be inaccurate. This was ostensibly in reaction to perceived advocacy masquerading as research within the educational literature. Leading in this effort has been the Institute for Educational Sciences (IES).

Upon the appointment of Grover “Russ” Whitehurst as the first director of the IES, only the gold standard (utilizing random controls and sophisticated analysis techniques to determine and measure the effect size) was the acceptable standard of knowledge. During the Whitehurst tenure, if a study wasn’t gold it wasn’t considered worthy of passing the agency’s standards and therefore not worthy of publication. To illustrate this, the published review standards of the time clearly stated that only studies with random assignment combined with low attrition would meet evidence standards. Studies with random assignment but with moderate or large attrition or studies without random assignment but with documented equivalence would meet evidence standards with reservations. All other studies would fail to meet evidence standards. (IES, 2008)

As a result of this demand for the gold standard, each of the educational laboratories has been tasked with the conduct of randomly controlled experiments or trials (RCTs) with sufficient samples to find effects of approximately 0.25 or greater. It should be acknowledged that while IES has not specified a minimum effect size for its funded studies, the level of its funding and the costs associated with the recruitment of multiple schools to participate in a randomly controlled experiment, typically results in approximately this level of power. Cost has been the primary limiting factor to the sample size. Of course, a limit on sample size also limits the power of a study. IES has also granted funding to numerous other researchers through its open competitions to conduct similar experiments, with like methodological requirements. As a result, more new data has begun to pour into the federal repository of such distilled knowledge, namely the What Works Clearinghouse http://ies.ed.gov/ncee/wwc/.

The term RCT refers to studies that randomly assign subjects to treatment or control conditions. In education this typically refers to the assignment of whole schools or classrooms to one condition or the other. Other potential sources of bias are statistically factored out or controlled. This results in rigorous evidence that the main effect of an innovation is the casual agent of the observed results. This is often referred to as the Gold Standard of research.

If we wish to compare and contrast the empirical practices of education with that of medicine, we must acknowledge the changes that have recently transpired in the field of medicine itself. Shortly after the publication of my earlier article, additional efforts to bring research findings to the medical practitioner emerged. Several agencies began to develop systematic reviews of existing medical research. Among these, and perhaps the most recognized and respected, is the Cochrane Collaboration (http://cochrane.org/cochrane-reviews) which began operation in 1993. A Cochrane review uses professional peer reviewed study protocols and standard definitions for evidentiary quality. Reviews such as this include multiple studies using methodologies to limit potential biases and random error found in single studies. (Chalmers et al 2002) The result is a distillation of the research base bearing on a particular medical question or topic for ready consumption of practicing medical personnel. The reviews are geared to produce actionable knowledge to inform and influence practice. The international popularity and widespread utilization of Cochrane reviews demonstrate its usefulness and the promise for such an approach. Other fields have developed similar resources for their practitioners. These include the Campbell Collaboration, which provides systematic reviews in the areas of crime and justice and social welfare as well as some limited reviews in education (http://www.campbellcollaboration.org/).

The educational equivalent of the Cochrane reviews has recently emerged and may be found in the What Works Clearinghouse. These are the educational practice guides. These are documents developed by nationally
recognized subject area specialists. This expert group reviews the evidentiary basis available on their subject. They then apply a set of rules for evidence to determine the certainty of the finding related to various instructional and programmatic practices. The official IES description of practice guides indicates that in order to meet a strong evidentiary requirement a practice must meet the standards of the agency and show consistent positive results. Moderate evidence means that a practice meets evidentiary standards and shows generally positive results. Low levels of evidence can result from expert opinion when lacking scientific evidence. (IES, 2011). Currently, there are fourteen practice guides found on the IES web site ranging from instruction for English language learners to drop out prevention. In a manner similar to medicine, several agencies have attempted to provide reviews to bridge the gap from educational research to practice. These include among others, the Center on Instruction (doing what works) http://www.centeroninstruction.org/, The Best Evidence Encyclopedia http://www.evestidence.org/ and the Promising Practices Network http://www.promisingpractices.net/.

It is further instructive to contrast the levels of development of professional research reviews and their levels of use by the two fields. The Cochrane reviews contain 4,500 reviews on pertinent medical topics. Despite this number of studies, it is only considered a good start by the Cochrane Consortium. The current web site states that they estimate that an additional 10,000 reviews are need to address all of the pressing medical questions or topics needing review. When we contrast this with the fourteen practice guides available to educators it is apparent that the level of available research in education lags well behind that of medicine and certainly the effort to provide distilled information to educational practitioners also is in its preliminary stages. In all likelihood this reflects the head start enjoyed by the field of medicine. Perhaps in the future we can expect to see much more activity in the production of educational practice guides as a resource for practitioners.

Two additional problems have emerged during the past decade. These have limited the usefulness of the educational data, as it applies to a medical model. There is the small, nagging reality that most common educational innovations seem to produce relatively small effects. They are most often insufficiently powered to produce an effect size of 0.25 or greater. This inconvenient truth combined with the limited funding of studies under the federal auspices, has produced numerous finding of non-significance. In retrospect it might have been much better to have funded fewer studies with a larger sample size than the larger number with less power. As a result we know about a lot of innovations that do not register on studies with limited power, and about a few innovations that do. In addition, the position taken historically by IES to discourage investigation into the contextual variables of program implementation has severely limited our ability to interpret the findings. The rate of positive finding in the recent batch of IES funded research shows only an occasional positive result.

The Power of a study is its ability to reliably find a difference if it exists of a certain size between the outcomes displayed by the experimental and control conditions. The size metric is often reported in standard deviation units. Thus we would say, the power of a study is .25 if it is determined that it has sufficient strength of design and sample size to find a difference of .25 parts of a standard deviation of difference between groups, if its exists, and the study can correctly identify this difference at least 95% of the time.

The second problem that emerged was the delay between the conceptualization of these kinds of studies and their conclusion. The average federally funded education RCT requires between three and four years to complete. While this time line is common in all other fields, the fact that in medicine the research pipeline is longer and more established tends to produce a dependable stream of research findings. Educators unused to these kinds of delayed studies have voiced the perception that study findings were forever being delayed. The resultant criticism of IES was well deserved. IES was described as producing information lacking relevance and timeliness. IES itself was described as being out of touch with the need for rapid, actionable information required by educators and school systems seeking to improve. New director John Easton has identified this as a concern. He has taken preliminary steps to address this second shortcoming by providing flexibility in the kinds of studies being funded and the methods approved by IES. For example, a regression-discontinuity method that uses a cutoff point rather than random assignment to form a comparison groups, is being recognized alongside the RCT methodology as constituting credible scientific evidence. In addition, IES has begun to recognize in certain circumstances single-case studies. Preliminary signs point to IES funding studies as much for the expected utility and usefulness to practitioners as for their rigor. Exactly how successful Easton and the agency will be in threading this needle toward rigorous yet timely and relevant results will soon be apparent. To date however, the knowledge base appears insufficient to fully support a medical model of education as envisioned in the earlier article.

An additional requirement for a medical model is to have widespread awareness and acceptance of the scientific research on program effectiveness. This must be true among educational practitioners, as well as the public. The evidence must be readily available. The expectation must emerge that available vetted evidence would be used to improve program or to guide the

Continued on next page
selection and implementation of programs. In the field of education this has just not yet happened. Obviously exceptions do exist, but the average educator remains blissfully under aware of the existence or potential usefulness of the What Works Clearinghouse. The existence of credible scientific evidence is the first and most obvious requirement of an empirically driven field.

In addition there are the equally important issues of awareness, access, acceptance and willingness to routinely incorporate into practice. There are real differences between the preparation of educators and medical personnel. There are real differences in the cultural norms that define each field. These are topics that deserve increased attention. The development of a culture of educators willing and eager to seek out and utilize the best data driven practices for improvement is worthy of its own focus. I would hope future papers on this topic are forthcoming in the professional literature.

The financial and organizational efforts expended by the federal government to create and collect credible scientific evidence of educational programmatic effectiveness, during the last decade, argue that the medical model has gained considerable traction. This vision remains a desirable goal of state and national policy makers. It may be that placing education on a similar empirical footing as that routinely employed in the field of medicine just takes more time and effort. Perhaps we are part way there. Only time will tell for certain.

References


-Bob Rayborn is a past WERA president (1996) and most recently was the director of the center for research, evaluation and assessment at Education Northwest in Portland, OR. Contact him at rayborn1@comcast.net.

Manuscript submitted January 18, 2011
Final revision submitted April 22, 2011
Accepted April 22, 2011
## Institute of Education Sciences Levels of Evidence for Practice Guides

<table>
<thead>
<tr>
<th>Levels</th>
<th>Operational Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strong</strong></td>
<td>In general, characterization of the evidence for a recommendation as strong requires both studies with high internal validity (i.e., studies whose designs can support causal conclusions) and studies with high external validity (i.e., studies that in total include enough of the range of participants and settings on which the recommendation is focused to support the conclusion that the results can be generalized to those participants and settings). Strong evidence for this practice guide is operationalized as:</td>
</tr>
<tr>
<td></td>
<td>• A systematic review of research that generally meets the standards of the What Works Clearinghouse (WWC) (see <a href="http://ies.ed.gov/ncee/wwc/">http://ies.ed.gov/ncee/wwc/</a>) and supports the effectiveness of a program, practice, or approach with no contradictory evidence of similar quality; OR</td>
</tr>
<tr>
<td></td>
<td>• Several well-designed, randomized controlled trials or well designed quasi-experiments that generally meet the WWC standards and support the effectiveness of a program, practice, or approach, with no contradictory evidence of similar quality; OR</td>
</tr>
<tr>
<td></td>
<td>• One large, well-designed, randomized controlled, multisite trial that meets the WWC standards and supports the effectiveness of a program, practice, or approach, with no contradictory evidence of similar quality; OR</td>
</tr>
<tr>
<td></td>
<td>• For assessments, evidence of reliability and validity that meets the Standards for Educational Psychological Testing (AERA, 1999).</td>
</tr>
<tr>
<td><strong>Moderate</strong></td>
<td>In general, characterization of the evidence for a recommendation as moderate requires studies with high internal validity but moderate external validity, or studies with higher external validity but moderate internal validity. In other words, moderate evidence is derived from studies that support strong causal conclusions but where generalization is uncertain, or studies that support the generality of a relationship but where the causality is uncertain. Moderate evidence for this practice guide is operationalized as:</td>
</tr>
<tr>
<td></td>
<td>• Experiments or quasi-experiments generally meeting the WWC standards and supporting the effectiveness of a program, practice, or approach with small sample sizes and/or other conditions of implementation or analysis that limit generalizability and no contrary evidence; OR</td>
</tr>
<tr>
<td></td>
<td>• Comparison group studies that do not demonstrate equivalence of groups at pretest and therefore do not meet the WWC standards but that (a) consistently show enhanced outcomes for participants experiencing a particular program, practice, or approach and (b) have no major flaws related to internal validity other than lack of demonstrated equivalence at pretest (e.g., only one teacher or one class per condition, unequal) amounts of instructional time, highly biased outcomes measures); OR</td>
</tr>
<tr>
<td></td>
<td>• Correlational research with strong statistical controls for selection bias and for discerning influence of endogenous factors and no contrary evidence; OR</td>
</tr>
<tr>
<td></td>
<td>• For assessments, evidence of reliability that meets the Standards for Educational and Psychological Testing (AERA, 1999) but with evidence of validity from samples not adequately representative of the population on which the recommendation is focused.</td>
</tr>
<tr>
<td><strong>Weak</strong></td>
<td>In general, characterization of the evidence for a recommendation as low means that the recommendation is based on expert opinion derived from strong findings or theories in related areas and/or expert opinion buttressed by direct evidence that does not rise to the moderate or strong levels. Low evidence is operationalized as evidence not meeting the standards for the moderate or high levels.</td>
</tr>
</tbody>
</table>

### Reference
Mathematics and Evidence-Based Practice: Using Clinical Research in a New Suit-of-Clothes
By Jessica Cohen, Ph.D., Jerry Johnson, Ph.D., and Kimberly Markworth, Ph.D.

Broadly defined, Evidence-Based Practice (EBP) is a systematic, reflective process for integrating relevant research evidence of medical expertise within a clinical environment. Two key factors affecting the success of EBP are access to current and past research and the professional ability to differentiate between high-quality and low-quality results. Despite the need and possible connections, little evidence exists regarding the translation of EBP to educational contexts. To overcome this void and perhaps a lack of knowledge, our purpose is to illustrate how EBP could be used for mathematics education purposes, using hypothetical scenarios, and identify the difficulties in implementing EBP in mathematics education.

An important first step is to recast EBP in a mathematics education context, namely the purposeful, explicit, and prudent use of current best evidence in guiding or making decisions about pedagogies, teaching strategies, assessments, curricular choices, or classroom interventions specific to the mathematics education of students. Translating “best evidence” from a medical context to an educational context is problematic, in part because of the difference in the type and availability of research. In the education context, “best evidence” implies access to scientifically-based research that provides statistically significant evidence of “treatments” as being positive “actions” for specific educational problems or concerns. In turn, EBP would include these five fundamental steps:

Step 1: Ask a well-formulated question specific to a mathematics education situation
Step 2: Identify articles and other evidence-based resources that offer viable responses to the question
Step 3: Critically appraise the evidence to assess its validity
Step 4: Apply the evidence as a response to the mathematics education situation
Step 5: Re-evaluate this evidence-based response and identify areas for improvement

For the purposes of this discussion, our focus will be on the use of EBP by curriculum specialists, administrators, and mathematics teachers.

Our literature review found little evidence specific to the use of EBP in a mathematics education context, let alone in an educational context. This is partially explained by alternative terminology for the educational use of clinical research. In 2001, the Federal mandate No Child Left Behind (NCLB) required schools to ensure their students had access to effective scientifically-based instructional strategies, defining “scientifically-based research” as “research that ...involves the application of rigorous, systematic, and objective procedures to obtain reliable and valid knowledge relevant to education activities and programs.” In 2005, the National Council of Teachers of Mathematics’ report, Harnessing the Power of Research for Practice, created “an integrated system to enable research to link with practice” and carefully defined both the meaning and interpretation of scientifically-based research within a mathematics education context.

Washington’s Office of Superintendent of Public Instruction has not addressed the use of EBP in schools, however it has directed attention towards using clinical research as part of the NCLB mandate. Finally, Montague & Dietz (2009) perhaps is the only journal article addressing the use of EBP in the teaching of mathematics, though its narrow focus is on the effects of using cognitive strategy instruction on the problem solving abilities of students with disabilities. This article also is an intentional assessment of the guidelines for using EBP as established in the Winter, 2005, issue of Exceptional Children.

Using scientific research to inform educational practice appears in literature, the work of professional organizations, and state and federal policies, but it rarely is in the specific context of EBP. Given this void, our task is to illustrate how EBP could be used for mathematics education purposes. Two specific examples will be provided, each illustrating a different role for EBP.

Example One
Consider the process a district uses to select a middle school mathematics curriculum, which currently has limited resemblance to the five-steps of EBP. This selection process is a ripe context for exploring the potential and limitations of applying EBP on the district level in mathematics education.

Step 1: What middle school mathematics curriculum will best meet the needs of the district’s student population and teachers?

Particular contextual factors need to be identified as part of this step, including – but not limited to – the make-up of the student population, prior mathematical experiences (i.e. elementary curricula), teacher expertise, and teachers’ comfort with various instructional strategies. Additional contextual factors may be considered, such as district parents’ attitudes towards the available mathematics curricula.

Step 2: Identify and collect research evidence that speaks to the middle school mathematics curricula under consideration.

Evidence from multiple resources should be considered, including the curriculum publishers, mathematics educators, and other school districts. In addition,
negative effects on test scores might be short-lived, or not considered, but judgment should be cautious; positive or curriculum and the mathematical ideas that are provided, and how students are transitioning to the new curriculum will best meet the needs of the school district? What prior experiences with mathematics were these students involved in that may have had an impact on their learning and engagement? What professional support was provided for curriculum implementation that may have affected teachers’ implementation of the curriculum? Application of these and other questions to the evidence that has been collected will assist those involved in selecting a curriculum. What sample of students and teachers were involved in this study? How does the sample match the student and teacher needs of our school district? What prior experiences with mathematics were these students involved in that may have had an impact on their learning and engagement? What professional support was provided for curriculum implementation that may have affected teachers’ implementation of the curriculum? Application of these and other questions to the evidence that has been gathered will assist those involved in selecting a few studies (perhaps one to three) that provide credible evidence of curricular effectiveness for the curricula that are under consideration.

Step 4: Applying the evidence. First, the evidence gathered in Step 3 must be applied to the question at hand: What middle school mathematics curriculum will best meet the needs of the school district? The information that was filtered in Step 3 is applied to the question identified in Step 1 in order to select and implement a new curriculum. Of course, curriculum implementation is lengthy, and it may be two years or more before the district has ample evidence to evaluate its effectiveness at the district level.

Step 5: Re-evaluating the application of evidence and identifying areas for improvement, which should begin earlier than this two-year lag. As the district evaluates teachers’ successes and difficulties with implementation of the new curriculum, what additional support for implementation might be provided, and how students are transitioning to the new curriculum and the mathematical ideas that are highlighted therein. Test scores will inevitably be considered, but judgment should be cautious; positive or negative effects on test scores might be short-lived, or not immediately apparent. Additionally, those involved with the curriculum selection process must stay up-to-date with the current literature, such that any additional evidence regarding the curriculum that has been chosen can be considered in light of the local evidence that is being gathered. Unfortunately, the necessary staff expertise and time constraints probably exceed a district’s willingness to undergo the EBP process, despite the useful information it provides.

Example Two Our literature search showed little research on the use of EBP in a classroom, especially in a mathematics classroom. However, it is possible to describe a hypothetical scenario in which a mathematics teacher uses EBP to make a teaching decision, and the process is shorter than that of selecting a curriculum.

Step 1: What are the most common student misconceptions about experimental probability and what activities help expose and correct them? Suppose a teacher is planning to teach a unit on the differences between experimental and theoretical probabilities. She has taught such a unit in the past and has anecdotal evidence that students struggle to understand the differences between the two, and are particularly challenged by cases when experimental probability differs from theoretical.

Step 2: Identifying articles and other resources to answer the question. For most teachers, these resources are difficult to access, particularly because school districts do not necessarily subscribe to databases of research journal articles. For mathematics teachers, regional journals, like Washington Mathematics, are often available; however, the scope of articles is limited and most are not grounded in research. Members of the National Council of Teachers of Mathematics have online access to practitioner journals, and possibly access to NCTM’s Journal for Research in Mathematics Education. Practitioner articles may be written from a teaching perspective, and every submission is reviewed by a group of three referees, at least one of whom is a university educator (possibly a researcher). The NCTM claims that each article submitted is considered from a research perspective. Moreover, as part of the editorial process, authors are often directed to include specific research as part of the revision process. Practitioner journals also include articles written by researchers designed to connect meaningful research to teaching practice, clearly benefitting teachers not trained to interpret, analyze, and apply research results into practice. Finally, few teachers will have access to research journals through university libraries in their community.

Step 3: Critically appraise the evidence. For classroom teachers this appraisal involves considering the source (both publication and author) of the material, comparing the results or arguments to personal experience and the experience of colleagues, and using experience and training in research analysis to evaluate results. Teacher preparation programs
often have minimal focus on evaluating and interpreting research, so this will prove to be a challenge to implementing EBP in the classroom. Yet, this is perhaps the most critical step for a practicing mathematics teacher, because the Internet provides an overwhelmingly large body of activities and writing with no editorial review, so teachers interested in using EBP must learn to cull quality resources out of all those available.

**Step 4: Applying the evidence.**

The teacher finds in the research that the idea of variability in experiments is challenging for students to understand, and she finds several classroom-tested activities in the practitioner articles which have been shown to help students build understanding of this idea. She decides to implement these activities in her classroom, and incorporates elements of assessments she found in the literature to use as an assessment tool to determine the effectiveness of her intervention.

**Step 5: Re-evaluating the evidence-based response and identify areas of improvement.**

In our hypothetical scenario, the teacher carefully observes students while working through the activities and analyzes the results on the post assessment. Although students seem to be successful in explaining and interpreting probabilities in the activities involving coin flips, some students struggle with similar questions on the post assessment in the context of dice. The teacher determines that she needs to include more experiments in different contexts or in situations where possible outcomes are broader, in order to improve students' understanding of variability in experimental probabilities.

**Summary**

We consider this article to be an exploratory discussion of how the five steps of EBP might be applied in a mathematics education context. Selecting a curriculum at any level is a lengthy, important, and expensive process, and teacher decisions directly impact students on a daily basis. Following the five steps of EBP could provide more coherence to the curriculum selection process and help justify and improve teaching decisions. An intended result in both examples is that those involved will be more prepared to both support and implement the decided actions.

As a structure for decision-making, EBP has some potential for application in mathematics education contexts. However, because mathematics education differs from clinical medicine, there are substantial obstacles to the implementation of EBP in mathematics education, which limit its viability. In addition to the expected time and financial constraints, key factors affecting the success of EBP are the availability of quality research in certain contexts, access to current and past research, the professional ability to differentiate between high-quality and low-quality results, and the adaptation or implementation of the research-based ideas into local situations. Its goal and systematic approach are commendable, but its application in mathematics education would be difficult at best at the present time.

The Winter, 2005, issue of Exceptional Children is helpful here in that it established guidelines for using EBP practices in an educational context. And, a final concern may be that of semantics, as EBP is a viable process but may appear different from the clinical research requirements established by federal and state agencies.

**Reference Notes**


--Jessica Cohen, Jerry Johnson, and Kimberly Markworth are mathematics department colleagues at Western Washington University in Bellingham. Jerry directs the delightful MathNEXUS Project, a partnership with secondary school mathematics teachers in local school districts. Each author conducts research in mathematics education. Contact them at jessica.cohen@wwu.edu, jerry.johnson@wwu.edu, and kimberly.markworth@wwu.edu.

Manuscript received March 9, 2011
Final revision received April 8, 2010
Accepted April 20, 2010
A Response to Mathematics and Evidence-Based Practice: Using Clinical Research in a New Suit-of-Clothes
By Brian Rick

As a former math teacher, the notion of applying EBP in math education interests me. This article illustrates how a clinical research process could be tailored to fit into the education world. So how good is the fit?

The familiar example of selecting a math curriculum easily transfers EBP from a clinical environment to an educational one, with clear and logical steps. The last step, however, re-evaluating the evidence, may be less familiar to us. Not everyone welcomes a second look after the huge commitment in new materials and professional development is made. I agree that limiting factors for EBP are time, money and access to relevant research. However, I think difficulty in differentiating between high- and low-quality research is a lesser concern.

The second example – using EBP to make a teaching decision – moves EBP onto a smaller scale. The fit of EBP in this scenario is seriously challenged by practicality. Of the limits noted above, the lack of time rises to the top. To give EBP a chance, I want to ask, “If I don’t have time to ask which are the most effective strategies to use, do I have time to use less-effective strategies?” For many teachers, we are content with reliance upon mentors and colleagues to garner ideas shared in lunchrooms, workshops and in journals such as this.

These two examples make me think that the relevance of EBP to educational questions depends greatly upon the nature of our research question and available resources.

So, does the suit fit? Well, I like the fabric... but for versatility I was looking for something a bit less dressy.

--Brian Rick is Assessment and Evaluation Specialist for Bellingham Public Schools. Brian is working on a searchable database for WERA publications. Contact him at Brian.Rick@bellinghamschools.org.

Making it Simple: Four Essentials to Measuring Impact
By Peter Hendrickson, Ph.D.

Why is it so difficult to determine if instructional innovations lead to improved student outcomes? We might guess that research focused on undergraduate science courses with higher education investigators would be both exhaustive and well designed. A recent paper found positive effect sizes but many problems with study design (Ruiz-Primo, Briggs, Iverson, and Shepherd, 2011).

The authors searched the professional literature around innovations in undergraduate science instruction to see which strategies made the greatest difference in course outcomes. The good news was that some student-centered instructional strategies seemed to be more powerful than others (e.g. collaborative learning). However, the litany of problems was extensive. Nearly half the 331 studies were rejected as summary statistics were not included to compute effect size. But the list of recommendations to avoid the design shortfalls was brief:

1. All studies should include descriptive statistics (sample sizes, means, standard deviations) for all treatment and control groups for each test event.
2. Random assignment to treatment or control is the most powerful study design.
3. If an RCT is not possible, researchers should demonstrate that the groups were comparable before the intervention, in particular for academic achievement.
4. Outcome measures need to be of valid and reliable.

Reference

--Editor
Evidence-Based Focus Resources for Special Education Programs
By Donnita Hawkins

Special educators wishing to make best use of research to guide their instruction face a confusing array of choices. This paper compares the support available from three organizations with an illustrative example. It could be argued that the individual approach in special education brings the practice closer to the evidence-based medical model than with other populations.

Background

Over time a shift has occurred in special education. Initial focus emphasized special while recent initiatives emphasize education (Spaulding, 2009). The reauthorization of the Individuals with Disabilities Education Act (IDEA) 2004 provided additional support for this shift by stating that special education students have “access to the general education curriculum in the regular classroom, to the maximum extent possible…” (20 U.S.C. § 1400(c)(5)(A), 2004). Students receiving special education services are expected to participate in the least restrictive environment to receive specially designed instruction. The No Child Left Behind Act (NCLB, 2001) sought to bolster this position maintaining;

Too often in the past, schools and LEAs have not expected students with disabilities to meet the same grade-level standards as other students. The NCLB Act sought to correct this problem by requiring each State to develop grade-level academic content and achievement standards that it expects all students—including students with disabilities—to meet (67 F.R. 71710, 71741).

This statement seeks to bridge IDEA and NCLB by focusing on access while increasing rigor and better defining general education curriculum for special education students. While this connection is imperative, special educators have struggled with providing access to curriculum and programs aligned to state standards. Several organizations have sought to assist with this issue by evaluating research and/or evidence-based practices. These organizations include:

- What Works Clearinghouse (http://ies.ed.gov/ncee/wwc/),
- Best Evidence Encyclopedia (www.bestevidence.org), and

Unfortunately, each organization has defined its own evaluation criteria with some being more stringent than others. Slavin (2008) aimed to assist educators by explaining the evaluation process each organization employs. He postulated that the What Works Clearinghouse utilizes the most the most rigorous study requirements and therefore may exclude research that does not meet their requirements. The National Center on Response to Intervention uses a less rigorous evaluation process and therefore includes more programs in their reporting. Table 1 provides a brief description of the review processes and an example of evaluation information reported for the Comparison of Findings - Intervention: Peer-Assisted Learning Strategies (PALS) for the three organizations.

Many special educators consult a 2006 EBP consensus statement from the Council on Exceptional Children (CEC) which does not restrict evidence to randomized controlled studies. However the route to adoption of EBP in medicine is by no means straight-forward. Umbarger (2007) noted reluctance on the part of physicians due to conflicting and out of date guidelines, outcomes expectations, lack of engagement with local practitioners, and suspicion about the motives behind the guidelines.

For the purposes of this paper, six special education directors were asked what process they are currently using to select programs and curriculum (Hawkins, 2011). The directors represent small, medium, and large districts from the east and west side of Washington. Responses indicate that the two districts implementing Response to Intervention (RTI) use the resources available from the National Center on Response to Intervention website as well as district educator expertise. The four districts that indicated they did not use RTI almost solely rely on internal review of materials and information provided from publishers. There is no formal review process, they stated.

Discussion

When special educators adopt a new program or curriculum, it is imperative that they be aware of these organizations and use the information provided. While this is time consuming, it is vital that special educators be critical consumers and take time to compile information and make informed decisions. It seems the biggest obstacle to selecting research or evidence-based practices is that as educators, we want an expedient approach so that we can obtain a curriculum and begin using it almost immediately. NCLB and IDEA require that educators, including special educators, be critical consumers.
### Table 1
**Review Process and Study Example Comparison**

<table>
<thead>
<tr>
<th>Organization</th>
<th>Review Processes</th>
<th>Comparison of Findings - Intervention: Peer-Assisted Learning Strategies (PALS) Information:</th>
</tr>
</thead>
<tbody>
<tr>
<td>What Works Clearinghouse</td>
<td>Stage 1: Determining the Relevance of a Study</td>
<td>Alphabetics- potentially positive effects- Avg. +19 percentile points</td>
</tr>
<tr>
<td></td>
<td>• Collect potentially relevant studies on the topic.</td>
<td>Fluency- potentially positive effects- Avg. +13 percentile points</td>
</tr>
<tr>
<td></td>
<td>• Consideration of timeframe, sampling, and reported findings.</td>
<td>Comprehension- potentially positive effects- Avg. +13 percentile points</td>
</tr>
<tr>
<td></td>
<td>Stage 2: Assessing the Strength of the Evidence that a Study Provides for the Intervention's Effectiveness</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Evaluate study design.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Consideration of evidence strength.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stage 3: Identifying Other Important Characteristics of a Study That Meets Evidence Standards (With or Without Reservations)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Evaluate effectiveness of interventions.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Utilizes meta-analyses where reviews must:</td>
<td>This site provides lengthy reports that require thorough reading.</td>
</tr>
<tr>
<td></td>
<td>1. Exhaustively seek out all relevant research.</td>
<td>PALS summation: Across 6 small studies of PALS, the weighted mean effect size was +0.44, and adding in the CWPT study, the mean for seven small studies of cooperative learning was +0.46.</td>
</tr>
<tr>
<td></td>
<td>2. Present quantitative summaries focusing on achievement outcomes.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. Focus on studies that used equivalent study groups.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. Summarize program outcomes in terms of effect sizes and statistical significance.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5. Focus on studies that lasted at least 12 weeks.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6. Avoid artificial laboratory studies.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Use control and experimental study measurements to avoid only analyzing studies</td>
<td></td>
</tr>
<tr>
<td></td>
<td>meant to measure a specific effect.</td>
<td></td>
</tr>
<tr>
<td>National Center on Response to Intervention</td>
<td>Step 1: Submission for Vendors</td>
<td>Classification accuracy- convincing evidence</td>
</tr>
<tr>
<td></td>
<td>Step 2: First- and Second-Level Review</td>
<td>Generalizability- moderate high</td>
</tr>
<tr>
<td></td>
<td>• Initially each reviewer reviews interventions individually.</td>
<td>Reliability- convincing evidence</td>
</tr>
<tr>
<td></td>
<td>• Both reviewers must come to agreement.</td>
<td>Validity- partially convincing evidence</td>
</tr>
<tr>
<td></td>
<td>Step 3: Interim Communication with Vendors</td>
<td>Disaggregated Reliability, Validity, and Classification Data for Diverse Populations- convincing evidence</td>
</tr>
<tr>
<td></td>
<td>• Summation of second-level review provided to vendors.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Step 4: Third-Level Review</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Additional evidence submitted by vendors reviewed.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Step 5: Finalization and Publication of Results</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Debrief among reviewers.</td>
<td></td>
</tr>
</tbody>
</table>

Continued on next page
References


--Donnita Hawkins is Special Education Coordinator, Secondary Education and School Improvement, OSPI. Contact her at donnita.hawkins@k12.wa.us.

Manuscript received February 15, 2011, Final Revision Received April 22, 2012 Accepted April 22, 2011.
Evidence Based Instruction: Tools & Processes for Observing & Advancing Effective Instructional Teaching Practices

By Jeannette LaFors, Ph.D. and Candace Gratama, Ed.D.

Introduction

What does an effective teacher do and how can observing instructional practices improve the overall instructional practice of a school faculty? While the field has identified “evidence-based” instructional elements, school leaders struggle with determining the extent to which these elements prevail and how they can ensure they do. This paper discusses instructional elements deemed “effective” and describes how tools and processes for observing these elements help schools in one case study district improve their instructional practice to support student learning.

Literature Review

Developing high quality instruction that improves learning outcomes for all students is a top priority for most schools and districts. Over the past two decades, research shows this is an ambitious task. Rigorous content standards, and the transformative changes in instructional practice associated with them, require substantial professional learning on the part of teachers (Cohen & Hill, 2000; Hubbard, Mehan, & Stein, 2006; Thompson & Zeuli, 1999); and the pressure on teachers has intensified under contemporary accountability policies (Finnigan & Gross, 2007; Valli & Buese, 2007).

To support this learning, school leaders have implemented walkthroughs. There are over 20 models of walkthroughs, varying in length of time, participants involved, and overarching purpose of the visit (for a range of models, see Ginsberg & Murphy, 2002, and The Center for Comprehensive School Reform and Improvement, 2007). One common thread ties them together: they all involve physically entering the classroom during instruction, taking notes, and using those notes to provide feedback to faculty and direct the course of future actions. These kinds of walkthroughs are explicitly characterized as “data-driven” and “non-judgmental,” offering a window for others to (a) view what actually happens behind classroom doors, and (b) use data to influence how resources like time, space, money, and talent are allocated. However, several enduring tensions – for instance, the natural human tendency to judge, a disregard for the complex nature of classroom interactions, and the strong professional norms of privacy in the teaching profession – have prevented walkthroughs from prompting substantial changes in instruction (Cohen, Raudenbush, & Ball, 2003; Lortie, 2002). We consider the success of the case study district to be due not only to its implementation efforts, but also to additional evidence-based practices linked to instructional improvement that reinforce the walkthrough processes.

Tools and Processes to Support Evidence-Based Practice

The STAR Classroom Observation Protocol™ from the BERC Group is a research instrument measuring the degree to which teachers employ reform-like teaching and learning ideals (adapted from How People Learn: Bridging Research and Practice, National Research Council, 1999). This type of teaching and learning correlates highly with student academic achievement (Abbott & Fouts, 2003; Baker, Gratama, & Peterson, 2010). The intent of the STAR Protocol is to measure what occurs in a given period of time for generalizing not to a single teacher, but to the school in the aggregate. In most cases, two to four observers visit each school for one day. Each observation lasts about 30 minutes, and observers calculate a score for each of the five Essential Components of the protocol and give each class session an Overall Component Score of 1 to 4. The STAR Process entails educators using the protocol to observe instruction, dialoguing with their colleagues, applying the discussion to their own practice, and reflecting on teaching and learning in their own classrooms.

The Classroom Walkthrough (CWT) developed by Teachscape, Inc. is a process designed to help educators identify instructional patterns and needs, and to monitor the progress of plans addressing them. By privileging evidence-based practices in its standard observation protocol, CWT helps educators focus on what matters for student achievement (e.g., Bloom et al, 1956; Kounin, 1970; Marzano, Pickering & Pollock, 2001; McTighe & O’Connor, 2005; Tomlinson, 1999). Furthermore, by helping teachers and leaders analyze and reflect on the instructional data collected, CWT supports a structure for planning and

Evidence-based strategies include: setting objectives, providing feedback, recognizing effort, asking questions, engaging students in cognitively demanding work, varying learning modalities and ensuring students have access to non-linguistic representation of ideas, helping students make connections to their lives, the real world and other subjects, and differentiating instruction to address the needs of individual learners (e.g., Bloom et al, 1956; Kounin, 1970; Marzano, Pickering, & Pollock, 2001, McTighe & O’Conner, 2005; Tomlinson, 1999).
monitoring instructional improvement (e.g., Darling-Hammond & McLaughlin, 1995; Dewey, 1933; Hord, 1997; McLaughlin and Talbert, 1993; Schön, 1983). The observation tool is designed to capture key practices in 4- to 7-minute observations, generating an aggregated profile of teaching and learning. From the initial profile, plans that focus on increasing evidence-based strategies can be monitored throughout the year in multiple cycles.

CWT and STAR are powerful supports for increasing the opportunities for students to experience effective instruction. To do so, both processes encourage discourse about quality teaching and learning, engage the school community to collect and analyze evidence, and support the development of action plans that can be evaluated through on-going data collection and analysis.

Case Study Findings

A small, growing district in Washington State with a predominantly Latino student population and nearly 80% of students qualifying for free- or reduced-priced lunches has demonstrated commitment to evidence-based practice and embraced both STAR and CWT processes over the past two and a half years as part of the Summit District Improvement Initiative funded by the state Office of the Superintendent of Instruction (OSPI). Measurement of instructional practice through STAR and CWT observations documents improvement in several key areas, and student achievement results are improving in math, a target area for professional development (Bates, 2011).

Fall 2008 baseline results on the STAR protocol observation study determined forty-two percent of classrooms observed lessons aligned to Powerful Teaching and Learning™ (PTL). After training 42 faculty members in STAR in 2008-2009 and 29 faculty in 2009-2010 to become facilitators, the district has worked diligently to support the STAR Protocol and Process. The most recent STAR protocol observation study results (Spring 2011) showed 63% of classrooms observed had lessons aligned to PTL. The district improved in each of the 5 components, with greatest gains in “Knowledge” (28 percentage-points) and similar gains of 9-10% for “Skills”, “Thinking”, “Application”, and “Relationships” (see Table 1).

<table>
<thead>
<tr>
<th>Components</th>
<th>Fall 2008</th>
<th>Spring 2010</th>
<th>Spring 2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skills</td>
<td>72%</td>
<td>76%</td>
<td>81%</td>
</tr>
<tr>
<td>Knowledge</td>
<td>37%</td>
<td>57%</td>
<td>65%</td>
</tr>
<tr>
<td>Thinking</td>
<td>36%</td>
<td>52%</td>
<td>46%</td>
</tr>
<tr>
<td>Application</td>
<td>24%</td>
<td>28%</td>
<td>9%</td>
</tr>
<tr>
<td>Relationships</td>
<td>84%</td>
<td>93%</td>
<td>10%</td>
</tr>
<tr>
<td>Overall</td>
<td>42%</td>
<td>56%</td>
<td>63%</td>
</tr>
</tbody>
</table>

These improvements in instructional quality align with improvements measured by CWT. Between Fall 2008 and Spring 2010, roughly 30 district leaders were trained in the CWT process and received over 12 days of professional development focused on specific strategies and how school leaders could support teachers’ understanding and adoption of the strategies. The figures below depict changes in how frequently specific focus strategies were observed as measured by the CWT classroom observation tool.

Additional evidence-based practices at the school and district level include those measured through the Educational Effectiveness Survey™ (EES). The figures are calculated mean scores on the staff survey for seven specific characteristics used in the Summit Initiative. The goal for each characteristic is a mean score of 4.0 (5.0 possible) at the end of 2010-2011. As depicted in Figure 3, the case study district improved from its baseline survey data in 2008 to Fall 2009 in all seven characteristics. Three critical areas, Clear and Shared Focus, Curriculum, Instruction, and Assessment, and District Characteristics met or exceeded the goal of 4.0.

Student Achievement

Despite the fact that the district lags behind the state in proficiency levels in math and reading in 12 out of 14 comparisons for 2010, comparing student achievement results from 2008 and 2010 shows case study district students grew in their performance on the state assessments in reading and math at faster rates than their peers across the state in 10 out of 14 comparisons grades 3 – 10 (See Table 2). In 2010, the district outperformed the state in math performance for two grades and came close to meeting the state performance for two grades in reading.

1 The units of analysis can be at both the district or school level (STAR & CWT) or with smaller communities of teachers such as grade levels or subject areas (CWT).
2 District and school administrators, classroom teachers, school coaches.
3 After the initial CWT training, the first strategies identified in 2008-2009 included: Setting Objectives, Providing Feedback, Providing Recognition and Recognizing Effort. In 2009-2010, professional development sessions included a focus on Summarizing & Note-taking (November); Cues/Questions/Advanced Organizers (January) and Cooperative Learning (March) with a focus on mathematics bolstered by a State-sponsored benchmark assessment series and technical assistance. Additional strategies were identified for 2010-2011: Homework & Practice (August); Similarities & Differences (September), and Generating and Testing Hypothesis (February).
Continued on next page
Fig. 3

Baseline 2008 vs. Fall 2009 district mean scores on the Educational Effectiveness Survey

Table 2


<table>
<thead>
<tr>
<th>Grade Level</th>
<th>District</th>
<th>State</th>
<th>District</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>4.7%</td>
<td>2.0%</td>
<td>7.7%</td>
<td>-9.9%</td>
</tr>
<tr>
<td>4</td>
<td>-13.2%</td>
<td>-7.4%</td>
<td>13.3%</td>
<td>0.2%</td>
</tr>
<tr>
<td>5</td>
<td>-14.4%</td>
<td>-7.9%</td>
<td>1.4%</td>
<td>-12.4%</td>
</tr>
<tr>
<td>6</td>
<td>-19.7%</td>
<td>-6.2%</td>
<td>-4.3%</td>
<td>5.7%</td>
</tr>
<tr>
<td>7</td>
<td>2.7%</td>
<td>0.5%</td>
<td>31.9%</td>
<td>9.5%</td>
</tr>
<tr>
<td>8</td>
<td>6.2%</td>
<td>4.7%</td>
<td>36.3%</td>
<td>-0.4%</td>
</tr>
<tr>
<td>10</td>
<td>0.5%</td>
<td>-3.5%</td>
<td>9.4%</td>
<td>-15.9%</td>
</tr>
</tbody>
</table>

1 This research-based instrument (Shannon & Bylsma, 2002) seeks to understand staff beliefs, attitudes, and practices known to exist in high performing schools by surveying the majority of staff in a school (all staff are provided the opportunity to respond to the survey). District faculty take the survey annually.

1 A mean score of 4.0 indicates an average response of “agree” across the staff.

Continued on next page
Conclusion & Further Investigation

The case study district maintained a clear focus on instruction as part of the Summit District Improvement Initiative and strongly implemented the CWT and STAR Protocol and Processes. Outcome data show improvements in staff members’ perceptions of school-based practices, in the implementation of specific focus strategies, and in Powerful Teaching and Learning. During the same period, state assessment results also increased in 10 of 14 areas in reading and math, with the greatest improvement in math. While these results are not causal, they do suggest there is a link between the support the district received through the initiative and improvement in assessment results. The impact of classroom evidence-based practices and school and district leadership evidence-based practices may be difficult to tease out from one another; but the district’s efforts to implement classroom walkthrough processes as they integrate and norm them into their district culture are worthy of further examination and possible replication.

References


--Jeannette LaFors is Senior Director of Professional Services at Teachscape, Inc. in San Francisco, CA, overseeing numerous projects and evaluating the impact of Teachscape products and services on instruction. She has been a teacher, instructional leader and educational researcher, and her research interests include instructional leadership, school improvement, college readiness and success, and ways to eliminate the achievement gap. Contact her at jilafors@teachcape.com.

--Candace Gratama is the Executive Vice President for The BERC Group in Bothell and serves as the primary investigator for multiple research and evaluation projects, and she has been central in the development of the STAR Classroom Observation Protocol. She has worked as a school counselor and is presently an adjunct professor at Seattle Pacific University. Her research interests include program evaluation, college readiness, student assessment, and school reform. Contact her at candace@bercgroup.com.

Manuscript received March 16, 2011
Final revision received April 22, 2011
Accepted for publication April 22, 2011
A Leadership Perspective on Implementing a STEM Initiative
By Margery Ginsberg, Ph.D., Chris Kinsey and Julia M. Zigarelli

Purpose, Background, and Context

The purpose of this article is to illuminate leadership perspectives on the implementation of a project-based, interdisciplinary approach to urban high school transformation using a science, technology, engineering, and math (STEM) focus. For several years, Cleveland High School has been one of Washington State’s lowest performing high schools. However, recent data suggest that the school is making progress towards overall improvement. In the last year, the attendance has increased from 84.6 percent to 89 percent (over 5%). On the Measure of Academic Progress (MAP) there has been a 7 percent increase in 9th grade reading and an 11 percent increase in 9th grade math. Further, there has been a 17 percent increase in 10th grade reading.

It is particularly significant to note that, although the national index for school improvement is 1.0, Cleveland High School has achieved a 1.5 index rating in 9th grade math and a 1.4 index rating in 9th grade reading.

As is the case with many schools across the United States, Cleveland High School has attracted significant attention from concerned stakeholders. As a consequence, school staff have initiated and adopted a range of school improvement approaches, with the most comprehensive change occurring in SY 2010-2011. During this time, Cleveland:

- shifted from a neighborhood school to a school drawing students from across the district,
- developed into two schools, each with a unique science, technology, engineering, and math (STEM) focus,
- incorporated a project-based interdisciplinary approach to teaching, and
- integrated the comprehensive use of technology.

This article examines lessons from recent changes, drawing on data from a 90-minute semi-structured interview with one of the school leaders and an author of this article, Assistant Principal Chris Kinsey. Given his permission and the school’s, we use Mr. Kinsey’s real name.

A semi-structured interview has a formalized, limited set of questions but is flexible, allowing new questions to emerge during the interview as a result of what the interviewee says. As a semi-structured interview, the interviewers had an initial set of questions that were intended to elicit leadership perspectives on transformation to a STEM school.

Methods

Data were collected by two graduate students and a University of Washington professor during a one and one-half hour interview with Chris Kinsey, one of two assistant principals. Each of the assistant principals leads one of the two themed schools at Cleveland. Mr. Kinsey leads “The School of Engineering and Design.” The interview questions focused on Mr. Kinsey’s experiences with and insights into the school’s recent transition process and were divided into three categories. These categories were 1) the experiences of teachers, administrators, and students at Cleveland High School, 2) the school’s preparation for the transition to a STEM school, and 3) challenges of school change.

The Broader Research Context

Although this article draws from one data source, the interview questions emerged as part of a broader study of school change focused on the experiences and perspectives of Cleveland High School students, teachers, and administrative leaders. The broader study involved 14 graduate students from the University of Washington-Seattle in cycles of action research. Acting as “participant-observers” and volunteers, graduate students met at Cleveland High School over a 10 week period. On-site participation allowed graduate students to visit classrooms on a regular basis, shadow students through their school day, serve as judges of interdisciplinary projects, convene focus groups of students and teachers, and interview administrative leaders.

These experiences provided the context for graduate students to learn about action research focused on school change in ways that held reciprocal value for the school. In other words, while graduate students had an authentic context within which to practice research methods and study issues of urban school renewal, Cleveland High School had a team of onsite “researchers” and volunteers to assist with their process of ongoing program improvement. For the interview with Mr. Kinsey, graduate students sought to probe more deeply into questions that arose for them as a result of their participation at Cleveland.

Professional Relationships and Theoretical Premises

In addition to providing context regarding the nature of the interview questions, it is also important to note the relationship that had been developed over time between Cleveland High School and the University of Washington.

Briefly, the UW professor had an existing relationship with the school which included assisting the school with developing a theoretical platform for launching and uniting a range of change initiatives. Given some of the historical tensions between on-the-ground educators and researchers from higher education who study the work of schools, these relationships provided a reasonably accessible context for reliable insights.

Continued on next page
The relationship between the UW professor who taught the action research course, and who is one of the authors of this article, and Cleveland High School evolved over four years, with three different principals, and two different superintendents. Two years prior to the interview, the professor co-founded The Center for Action, Inquiry, and Motivation (AIM) to provide on-site support for five components of school change that are associated with significant school improvement on multiple indicators of effectiveness, such as test scores, attendance rates, discipline referrals, and so forth. These components are:

- a shared pedagogical language
- multiple approaches to instructional collaboration
- routine use of data to inform instructional decisions
- a strong team of teacher leaders and stakeholder-advocates
- a school identity or signature about which other schools seek to learn.

The school had been working with the professor to develop these five components as a platform for change.

Data Analysis

Data analysis was an iterative process that began with data collection. Notes were made during and at the end of the interviews. These notes included quotes, descriptions, and impressions of the information that Mr. Kinsey shared. Notes helped to shape subsequent interview questions. Although data were not recorded, researchers compared their notes and impressions for accuracy immediately after the interview. Data were then coded to identify recurring themes and concepts. Some codes, such as “preparation” and “challenges” were created prior to the categorizing stage of data analysis based on experiences, relevant literature, and the study’s primary research questions. Other codes, such as “a focus on students”, “the importance of relationships and transparency,” “the importance of reflection”, and “new ways of measuring success” emerged from the process of reading and rereading the transcribed interviews.

Notes from the interview, informed by literature on school change (Seashore-Lewis, 2010, Fullan, 2002, Elmore, 2007) were coded, categorized, reviewed and summarized in narrative text to articulate emergent findings. An outline of research questions and raw interview data that address the primary purposes of this interview are provided in “Exhibit 1.1.” (These notes were provided courtesy of UW graduate student, Julia Warth). Given the brevity of this article, raw data are provided as an opportunity for readers to engage in sense making related to their own experiences, understanding of literature, and research interests (see Appendix).

Discussion

The analysis we provide is preliminary. However, one broad category of concern has emerged and it has been corroborated by multiple reviewers. It is consistent with research on major change initiatives (Louis & Gordon, 2006; City, et. al, 2009) and it resonates with administrative leaders and teachers. This theme, which we entitle, Identify and Communicate Strategies to Work Effectively with High Levels of Stress, has three primary subcategories: be transparent, maintain the centrality of students’ learning needs, and encourage collaboration and reflection.

Identify and Communicate Strategies to Work Effectively with High Levels of Stress

Even under ordinary conditions, the need to effectively manage stress within the school community is a significant aspect of educational leadership (Fullan, 2002; Goldberg, 2006). An everyday responsibility of school based leaders is to simultaneously buffer the ebb and flow of a host of policy impacts on students, families, and members of the broader community. Complex change initiatives such as Cleveland High School’s transformation to a STEM school, pose a particular challenge. “How-to” manuals and professional development seminars are often insufficient because local contexts differ and a vision of “success” is a work in progress. At Cleveland High School, stress on all stakeholders was exacerbated by pressure to rapidly restructure, responsibility for defining uncharted territory regarding STEM implementation in this particular district and school, accountability for providing evidence of effectiveness that represents and encourages student (and teacher) learning and motivation, and need for reliable and predictable resources to support ongoing professional growth and development.

When asked what Mr. Kinsey would like others to understand from this interview, without pausing he stated, “Change is hard!” “Leaders need to make sure that other people understand that a school is changing in the right direction.”

Be transparent

A theme that Mr. Kinsey frequently emphasized was transparency. Interview data suggest that transparency relates to communicating expectations and listening well. Yet transparency, alone, can contribute to widespread insecurity and concern. Interview data suggests that lessons regarding transparency include:

- Combine expectations with discussions about resources and support
- Understand that transparency requires two-way listening and communication
- Emphasize aspects of existing and widely agreed upon priorities that will continue.

At Cleveland, when leaders announced that change or closure was inevitable, they simultaneously discussed resources and activities that would assist teachers in
developing the skills needed for project-based learning and technology integration. This was also an approach they used in communication with students. Although many students initially responded, “hell no - more math and going to school longer than anyone else in the district?” they were also intrigued by one to one technology and an education that was considered to be “cutting edge.”

Leaders were also clear about the ways in which new skills could be united with the school’s existing instructional framework of the 4R’s (relationships, relevance, rigor, and results). On several occasions, leaders referenced and demonstrated how project-based learning could be designed, implemented, and improved with the 4R’s serving as a pedagogical compass.

Further, leaders openly acknowledged challenges, including the challenges of implementing change that is instructional as well as structural. They consistently engaged with teachers in conversations that led to an understanding that “...in many ways implementation could feel like student-teaching all over again,” and “project-based learning would be a major pedagogical shift—teachers would need to move away from being head of all knowledge in the classroom.”

Interview data also suggests the need for transparent two-way communication. According to Mr. Kinsey, leaders tried “to listen in-depth, by watching people’s actions, not just what they said.” To respond to teachers’ concern that they would need to put in 70-80 hours a week but would not receive fiscal compensation, leaders were vigilant about using time well. During the implementation year, planning time for teachers was protected to encourage collaboration on project development. Upon reflection, Mr. Kinsey noted that while this was essential, there were additional needs that surfaced for which time had not been as clearly brokered. For example, teachers became aware of their need to move fully understand “…how to manage laptop use...” and to teach students to say, “Here is what I need to know.” Further, home visits, a valued initiative to enhance communication and cultural competence, were sidelined to make room for significant competing priorities.

**Maintain a focus on students**

A second prominent theme that presented itself early in the interview and remained constant throughout was the importance of maintaining a focus on students. As previously mentioned, at Cleveland High School this included the need to continuously increase work toward cultural competence—given that “most teachers don’t live in the community and struggle in the same way the families do.”

To address this, leaders asked teachers to continue work they had been doing over the last two years – setting professional goals in ways that connected to the progress of four very different kinds of learners. At Cleveland, each teacher maintains a sharp focus on student learning and instructional improvement by becoming particularly aware of the lives, interests, and academic strengths of four different learners. At the beginning of the year, leaders work with each teacher to select two low-performing students, as well as a middle- and a high-performing student, each of who could serve as a touchstone for strengthening instructional practice.

According to Mr. Kinsey, “When teachers were asked to pick four students, it forced people to build relationships. Teachers had to look at reasons for struggle, for getting by, and for success to help plan lessons that would help students succeed... It created intentional planning that pushes every student.” In many ways following the progress of four students reinforced the connection between professional development and accountability as reciprocal processes. This idea is substantiated by several theorists, among them Richard Elmore. Elmore (1996) reiterates what Mr. Kinsey has learned from experience and has been working with colleagues to confront “…most educational reforms never reach, much less influence, long standing patterns of teaching practice, and are therefore largely pointless if their intention is to improve student learning...when schools seem to be constantly changing, teaching practice changes so little and on so small a scale” (p.6). He adds, “The core of schooling, defined in how teachers relate to students around knowledge, how teachers relate to other teachers in the course of their daily work, how students are grouped for purposes instruction, how content is allocated to time, and how students' work is assessed – changes very little. The changes that do tend to stick are those most distant from the core” (p. 7).

**Encourage collaboration and reflective practice among educators**

A third theme that is evident through the data is the importance of collaboration and reflection, or collaborative reflection, on the learning process. These practices provide a way for educators to access shared knowledge, reassess assumptions, and ask better questions about improving instruction. At Cleveland High School collaboration had several purposes. It served as a reminder of “... the power of the group coming together.” It provided a context to “build the program from within the school and from the ground up.” Finally, it allowed teachers to share and vet projects in professional learning communities. This may be one of the reasons that “all teachers are now doing projects,” and “…student presentations are announced to all staff to create authentic experiences for kids to present.”

**Future Research Directions**

The analysis of data is preliminary. Further, the themes from a single interview, however rich, require corroboration. In addition to interviews with other educational leaders at Cleveland and beyond, a comprehensive understanding of change requires perspectives from and research partnerships with
teachers, students, family members, district staff, and community partners. Catherine Lewis (2006) refers to this as “local proof.” As the co-authoring of this article with Mr. Kinsey suggests, local research partnerships with K-12 educators provide an opportunity to more accurately represent and intellectually probe the real work of committed educators who are doing it.

Appendix

1. How did you get to this point in leadership and work in your career?
   - This is my 11th or 12th year with the Seattle School District
   - Called by district to go to Cleveland High School with Princess Shareef
     - School needed a leadership change
     - This is my 3rd year here
     - The first year we focused on instruction and were also placed on the school closure/transformation list
     - The second year was spent planning STEM transition

2. What would you like to be sure to communicate through this interview?
   - Change takes time and is hard!
   - I want to communicate change is moving in the right direction.

3. For the past three years, teachers have selected and followed the progress of four students who were touchstones for differentiating curriculum. The sample included two low-performing students, a middle performing student, and an academically advanced student. Describe why you thought the idea of following four students would be a productive way for faculty to learn and how this may influence your current understanding of effective implementation of STEM.
   - Following 4 students emerged from our home visits. Each teacher was asked to visit the home of 4-5 students.
   - Teachers had to look at reasons for struggle, for getting by, and for success to help plan lessons that would help students succeed
   - Two students who were struggling were chosen by each teacher because one might drop out
   - Another reason home visits were important is because they moved us beyond classroom instruction and required the adults to make themselves vulnerable as well
   - When you look at the commonalities between kids who struggle and the struggles of successful students you take into account everyone’s strengths
   - Creates intentional planning that pushes every student
   - Increase cultural competence—teachers don’t live in the community and struggle in the same way the families do

4. As you think about products and conversations related to teachers following four students, what is an outstanding example?
   - We have strengthened common planning time for teachers and are using exit tickets with reflective questions to continue conversations
     - Always thinking about who’s struggling and are we pushing the high-achievers
   - The higher-achieving kids are not on track to make yearly progress—we need to push everyone
   - Reflective writing has been part of the process. It is personal and authentic—the collaboration comes when you share it out
   - There have been some amazing portfolios detailing how following four students has changed teachers as educators

5. When you look back at last year’s professional development related to preparing for implementing STEM, what aspects of it were particularly significant? What would you do to ensure even greater support?
   - We were very transparent—this is where we are going and this is the professional development to go along with it
   - The week before school started we focused on “what was it like to be in high school?”
     - It was a week of writing and modeling instructional practices
     - Math, science, and social studies teachers prepared lessons that teachers were “students” in—revealed expertise—powerful that it was teachers themselves, not outsiders coming in to model
   - Preparation included not only STEM but full-inclusion—special education kids in general education classes, getting the same curricula, and English learners in general education classes, as well.
   - Project-based learning was a major pedagogical shift—Teachers had to move away from being head of all knowledge in the classroom
   - We also needed more support for kids who had

Continued on next page
not learned this way before.

• We had quite a bit of externally imposed professional development because of STEM focus.

6. What have been some of the greatest challenges to implementing STEM? What have been the greatest successes?

• It is important to acknowledge the challenges—wanted to walk away, which has never happened before
• Selling it to the staff, students, and community was a challenge
• We asked staff to completely change the way that they teach and how they are evaluated, but we didn’t know what that looked like at the time
• When something like this becomes public, you see how much work is needed on cultural competence
• Developing the program was a challenge—what does it look like? There is no curriculum out there, built from the ground up
• Financing and the master schedule a challenge
• Politics—board, district office—school had to act like a buffer, administration shielded teachers from the distractions of politics, but still had to give opportunity to plan and advocate for school
• Finding time to coordinate
• Implementation year was the hardest, like student teaching all over again
• How do you ask someone to put in 70-80 hours a week but not get paid and make the time seem valuable?

7. Successes

• Last summer the entire staff went to a week-long institute in Indiana. Taking staff to Indianapolis was a success
  o There was power of the group coming together
• The number of people that opted in with blindfolds on—really wanted this to happen
  o Building program from within the school and from the ground up
• All teachers doing projects now (math is primarily problem-based rather than project-based)
  o We share projects in professional learning community, get together and plan—would like to see teachers vet projects more, but trust them to use the time for what is most effective for them
  o Student presentations are announced to all staff to create authentic experiences for kids to present

8. When implementing STEM, what was the response among 1) teachers, 2) students, and 3) the community?

• The initial teacher response: We gathered all the teachers in the auditorium before the board meeting to tell them what will be happening—either closing or transform
• Rollercoaster for teachers—allowed leaders to listen in-depth, by watching people’s actions, not just what they said
• Student response: hell no, more math and going to school longer than anyone else in the district
• Started to sell to students—one to one technology, yes we’ll have more school, but education is cutting edge
• Community response: Another change? Want to do it all again?
  o When the new Cleveland was built, was designed to have four small academies—didn’t work, how will this be different?
  o Make an opportunity for kids in the South end, not for North end kids who can’t get into another prestigious high school.
  o Community invited to be involved in planning—parents and partners
• Still a disconnect between what the industry wants from the high school and what colleges want—part of professional development
• Cleveland is an option school, not a neighborhood school
  o Had to sell the school at every middle school PTA meeting
  o No one was assigned to the school, kids had to choose to enroll

9. What numeric evidence or qualitative anecdotal evidence do you have to show that the teaching and learning are improving? What kind of evidence would you like to collect in the future?

• 9th grade attendance rate is higher
• Anecdotal success—kids/culture, doing projects, working in halls
• MAP scores—need to push higher achieving kids more, but doing well with lower-achieving kids
• Would like kids to keep bi-monthly reflections/journal
• State tests keep changing
  o Need 74 and 70 students to pass writing and reading, respectively, to meet improvement goals
• District climate survey at the end of the year—want one for the beginning and the middle of the year
• Go back to home visits—need more hard questions, attendance, grades, not just open-ended
• Vision and mission for school/STEM is more clear
• Get staff and students to write a personal
manifesto related to the vision and mission at the beginning of the year and at the end

• Measure kids by presentations—afraid to speak at the beginning of the year, now can’t get them to stop at the time limit. The documentation and comments of project judges from the broader community provide evidence of students’ strong performance and learning.

• We have insights from community members who help us be transparent and reflective – who to come to Cleveland and see what we are doing

• Getting more real
  o Business and industry: here is what you are learning and how we use it out in the real world
  o Same with higher education
  o How to collaborate, think critically
  o Show teachers and students how this is applied

10. If you could alter one thing about the way in which STEM was implemented at Cleveland, or were giving advice on STEM implementation for a future school, what would it be?

• Be authentic!
  o Create meaningful change for community and staff
• Keep students at forefront of planning
• Regarding Change-
  o Would have liked to have started at just 9th grade and rolled up
  o Should we have rolled it out throughout the year?
  o We were exhausted by winter break
  o Admin consistently asks, How could we have better supported implementation?
  o Need to evaluate initiative support
• Did not anticipate the need for some professional development
  o How do you manage laptop use?
  o How do we scaffold adult learning?
  o How do you teach a kid to say, “Here is what I need to know”?

11. Are there any other thoughts you would like to share?

• Would like the acronym to be STEAM, the arts are taking a hit with the focus on STEM (With this in mind, Dr. Ginsberg would like us to be STEM – Significant Teaching for Equity and Motivation).
  o Studies show that arts students are college-bound, need to help them too
• We need to continue our emphasis on the four Rs - Relationships, Relevant, Rigor, and Results - and ensure that equity is a core value in what we do here
  o The 4R’s may not come up in every conversation, but teachers will be able to talk about in every lesson plan
  o Personally, I ask: How can I find things that are refreshing and revitalizing?
  o Want to be with kids, not in the office
• Have to surround yourself with the very best, we have done that
• Lucky to have a great group of kids—they chose to come here

References

--Margery Ginsberg is an associate professor in the Department of Educational Leadership and Policy Studies at the University of Washington. She is a frequent contributor to WERA activities. Contact her at gumsbm@u.washington.edu.

Chris Kinney has worked in Seattle Schools for 11 years, three of which have been as assistant principal at Cleveland High School. He was Dean of Academic Interventions at Nathan Hale High School, House Administrator at Eckstein Middle School, and a 7th grade Language Arts and Social Studies teacher at Meany Middle School. Contact him at chkinsey@seattleschools.org.

Julia Zigarelli is a UW Research Associate pursuing a Masters of Education in School Psychology. Her interests concern the effects of culturally responsive teaching on student outcomes, as well as adult motivation in educational practice. Contact her at juliazig@uw.edu.

Manuscript received April 3, 2011
Final revision received April 20, 2011
Accepted April 21, 2011
School Conditional Growth Model: How to Make the “Apples to Apples” Comparison Possible?
By Yun Xiang, Ph.D. and Carl Hauser, Ph.D.

Introduction

Many believe that schools with a large proportion of disadvantaged students may be unfairly judged when they fail to make Adequate Yearly Progress (AYP). One option for evaluating the extent to which this is the case is a growth model that emphasizes individual student progress. Growth models allow contextual characteristics such as prior achievement to be taken into account. Compared with the simple regression with covariates, the model used in this study contains a few features. It first takes into account the nesting structure of educational data, which, in this case, is that scores are nested within students and students are nested within schools. Second, it looks at growth in a longitudinal way instead of simple two-time gain scores. Unlike gain scores, student or school growth trajectory is a more accurate reflection of achievement progress based on multiple time points.

When school growth rate is used to evaluate school effectiveness, the growth model relies on an important assumption; specifically, school growth is not confounded with school characteristics. For example, student growth rates in schools with large proportions of disadvantaged students are assumed to be similar to the rates in other schools. Is this a valid assumption? Growth-based accountability systems need to be subjected to scrutiny.

To help account for aspects of school environment in the evaluation and comparison of schools, a conditional growth model is used. Conditional growth models can be useful tools for examining and understanding the interrelated social demands of learning. In this case we are addressing the possibility that schools with a large proportion of disadvantaged students may be unfairly evaluated and judged, we take into account school (control for) characteristics in the conditional growth model and develop conditional school growth to make the so-called “apples to apples” comparison. Unconditional growth models depict school growth that depends solely on time, while conditional growth models explicitly account for school characteristics.

Method

Data and Measurement

The study includes test records from almost 50,000 students in 476 schools located in a single state. Instead of randomly selecting schools, we purposely included every school that has used the MAP test in this state to demonstrate how to look at school growth within a certain context (in one state or one district). The study focuses on one cohort of students by tracking their mathematic achievement growth from Grade 3 in term Fall 2006 to Grade 5 in term Spring 2009. The data came from the Growth Research Database (GRD™) developed and maintained by the Northwest Evaluation Association (NWEA).

One important criterion to conduct such a growth study is that the measurement scales remain stable over time. Hence, the skills possessed by a student with a score of 210 correspond to the skills processed by another student with a score of 210, even though the student may be at different points in their education (e.g. different grades). It is that constancy of measurement that growth to be measured and student performance to be compared across time.

Methods and Procedures

To use achievement growth trajectories to evaluate schools, we conducted a three-step analysis:
The first step, the unconditional growth model, examines the general trends in school mean growth. The unconditional growth provides insight into how

• students in each school change over time in general. The models are unconditional in the sense that growth is modeled as a function of time; common school predictors, such as percent Free-Reduced Lunch (FRL) students, percent minority students, school size, etc, are not used. The schools are ranked based on their estimated mean growth rates out of 476 schools.

• The second step, the conditional growth model, investigates school growth rate on school-level attributes. In this model, we focus on what school contextual characteristics are associated with school growth rate.

• The third step, the conditional latent regression model, examines the relationship between school growth rate and school initial status. When initial status is employed as a predictor of rate of change in the model, the regression is termed as a latent variable regression (Raudenbush & Bryk, 2002; Seltzer, Choi, & Thum, 2003).

Continued on next page
In the end, we discuss how schools can be fairly evaluated when their school characteristics are taken into account. Then we rank schools based on estimated mean growth rates after controlling for selected school contextual variables and school mean initial status. The final model we used in this study is called the conditional latent regression growth model. The latent regression feature considers school initial achievement in the model while the conditional feature considers school contextual characteristics.

**Results**

Based on the unconditional model and the conditional latent regression model, we ranked all 476 schools in one state, and then compared the difference between the two rankings. The correlation between the two sets of rankings based on the two models is very high ($r=0.99$). It is anticipated that the two models in general do not produce dramatically different results. However, when we looked at individual schools, a few interesting findings were revealed:

- With unconditional versus conditional latent regression models, one-third of schools ranked differently by more than 4 positions, 10% of schools ranked differently by more than 10 positions, 5% of schools ranked differently by more than 20 positions.  

- With unconditional versus conditional latent regression models, smaller schools, schools with larger proportions of disadvantaged students, and schools with fewer resources tended to be ranked differently by the unconditional and the conditional models.

The second finding is particularly thought provoking. We often hear questions from school districts: We are a school with a large group of minority students, or a school with limited resources, or a small school in rural area. Can we be compared to schools that are similar to us? Our study shows that the request for an apples-to-apples comparison needs to be addressed since disadvantaged schools tend to rank differently in a model that considers school contexts.

What is the implication for individual schools if schools are evaluated solely by their mean growth rates as opposed to growth rates that are conditioned on school characteristics? To illustrate this question, we randomly selected 18 schools out of 468 schools in our sample and ranked them based on their mean rate of change in two models (see Figures 1 & 2). Each bar represents a school’s mean rate of change with the upper line and the lower line representing the 95% confidence interval around the true mean.

First, we found that three schools (S10, S11 and S16) changed rankings based on their mean rates of change. However, when we also look at their standard error of estimation, we found that the confidence intervals for school no. 10 and school no. 11 mostly overlap. It indicates that even when they switch positions based on different models, the change is not significant. Their growth rates do not significantly differ from each other. School no. 16 is another story. As Figure 2 shows, this school did change its ranking since there was no overlap between its 95% confidence interval and school no. 17’s. It tells us that based on the unconditional model (see Figure 1), school no. 16 had a lower ranking than school no. 17. However, after considering school contextual characteristics and initial achievement, this school ranked higher than school no. 17 (see Figure 2).

Another notable finding was that the corresponding confidence intervals were larger for the conditional model (see Figures 1 and 2). This occurs because the hierarchical linear growth model is based on large-sample theory. When a school has a small size, the latent variable regression coefficient in the conditional model reflects more uncertainty about all unknowns in the model (Seltzer, Wong, & Bryk, 1996). That is why schools with a small student sample tend to rank differently by the two models. It implies that the latent regression conditional model will be more sensitive to the cohort size in a school.
Discussion

The study is a demonstration of how schools can be evaluated in the context of a school accountability system. It has become common for schools to be evaluated based on a two-dimension matrix recommended in a Council of Chief State School Officers (2005) report. *Policymakers' Guide to Growth Models for School Accountability: How Do Accountability Models Differ?* This matrix is presented in Figure 3 which captures growth and status. Schools in Group IV produce both high growth and high status and schools in Group I will be identified as schools needing to improve. There are also many schools that have mixed results (in Groups II and III). Only looking at one dimension of the matrix, either status or growth can result in a misleading conclusion about school performance.

![Figure 2: School ranking based on the conditional model.](image)

![Figure 3: Two dimensions matrix of school accountability models (CCSSO, 2005)](image)

When researchers and policy makers recommend a two-dimension matrix of initial score by growth rate as a basis for evaluating schools, this study suggests that the inclusion of school characteristics could provide additional useful information.

Figure 4 is an example of expanding the matrix of initial score by rate of change by adding relevant school factors, in our example, percent of students eligible for free and reduced lunch (FRL). The figure...
shows that schools with a higher percentage of FRL students tend to fall in Quadrant 3—low initial status with low growth. Schools with low percentage of FRL students mostly fall in Quadrant 1—high initial status with high growth. Without flagging schools as high, medium or low FRL percentage, we would not know the different patterns of schools’ performance and growth when their FRL percentage varied, and more importantly, we would not know how some high-poverty schools did an outstanding job in maintaining their students growth (see red dots at the very top of Quadrant 4). It could be misleading to simply evaluate schools based on their mean rate of change without considering relevant student or school characteristics. Figure 4 shows that considering growth alone, is likely to omit important correlates of growth that would otherwise help us to make more appropriate apples-to-apples comparison between schools. **Looking at student achievement growth is one step forward, but growth cannot be used as a solitary measure of school effectiveness.**

![Full Model - Free/Reduced Lunch Percentage](image)

*Figure 4
Matrix of initial score by rate of change with Free/Reduced Lunch Percentage*

It is not our intention to show what specific school characteristics should be considered in school evaluation. We recommend that researchers and administrators explore how schools are different from each other in their system. For example, in one district, schools may differ greatly as to their socio-economic status; in another district, schools may only differ greatly as to the level of students’ English language proficiency. Another cautionary note is that we are not suggesting that the current accountability system should be replaced by the conditional growth model we presented in this article. The current system emphasizes that every student ought to have an opportunity to reach a certain standard and schools are held accountable for that. Our approach focuses more on how to evaluate or compare schools based on the factors such as their previous performance, what characteristics they have, and how they grow differently.

**References**


--Yun Xiang is a Research Specialist at the Kingsbury Center at Northwest Evaluation Association in Portland, OR. Yun’s areas of expertise include growth modeling, program evaluation, and handling of large datasets. Her recent published works and presentations include an investigation of achievement gaps and summer loss, applications of growth modeling in evaluating schools and addressing student mobility, an analysis of school productivity by comparing different value-added models, and research involving the professional satisfaction of teachers. Contact her at yun.xiang@nwea.org.

--Carl Hauser is a Senior Research Specialist, Measurement and Data Service at Northwest Evaluation Association. His current interests include test item drift, differential items functioning in computerized adaptive tests, the 2011 version of the RIT Scale Norm Study, using student item response time as a cautionary variable in considering test score validity, diagnostic skills assessment, and documenting the technical characteristics of computerized adaptive tests. Carl was formerly a director of research and evaluation in the Olympia School District. Contact him at carl.hauser@nwea.org.

Manuscript submitted March 2, 2011
Final revision April 6, 2011
Accepted April 8, 2011
The Mathematics Benchmark Assessments (MBAs) are interim assessments developed to assess the K-8, algebra 1, geometry, algebra 2 and math 1 – 3 WA state mathematics standards. The MBAs are administered in the fall, winter and spring to provide ongoing information about the teaching and learning of standards. In 2010-2011 school year, the MBAs were implemented through the OSPI Washington Improvement and Implementation Network (WIIN) Center in 22 school districts serving over 95,000 students.

The MBAs were first implemented in eight school districts participating in the Summit Improvement Initiative during the 2009-2010 school year. An analysis of the spring 2010 MSP results revealed that the growth in mathematics achievement in these eight school districts was higher than the state in 40 of the 56 grades tested.

Table 1a

Number of Summit school districts with positive, negative or neutral change in spring 2010 math MSP/HSPE results

<table>
<thead>
<tr>
<th>Grade</th>
<th>Positive Change</th>
<th>Negative Change</th>
<th>Neutral Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 3</td>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Grade 4</td>
<td>5</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Grade 5</td>
<td>3</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Grade 6</td>
<td>7</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Grade 7</td>
<td>6</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Grade 8</td>
<td>6</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Grade 10</td>
<td>0</td>
<td>7</td>
<td>1</td>
</tr>
</tbody>
</table>

Two school districts – Othello and Sunnyside - saw growth in every grade level, 3 through 8, and one school district – Tukwila – saw growth in all but one grade level, 3 through 8. These three school districts also serve the highest ELL populations of the Summit districts. In every instance of positive growth, the growth outpaced the state (Bates, 2011).

Clearly this growth can not only be attributed to the MBA implementation. The Summit District Improvement Initiative includes many improvement supports, processes and tools. From comprehensive district action planning to data analysis services and classroom walkthrough tools, these districts receive intensive professional development and support from five external partners. The majority of this support is content neutral. However, similar growth was not realized in reading and only one thing was different in mathematics: the implementation and support of the MBAs.

Several implementation patterns were observed in the Summit districts and schools that achieved the most growth. Most notably, these districts had strong leadership in place that expected and supported principals and teachers to analyze the data after each MBA and use the data to make improvements. For example, principals and instructional coaches ensured all teachers had access to the MBA data and had a structured time for data review. The district leadership was positive about the MBA process. Although there were some errors in the items and data was delayed in getting to schools, leaders highlighted its usefulness and the learning that could still occur in its use. Further, these districts took full advantage of the OSPI professional development and technical assistance and ensured its content and important information was communicated to all appropriate staff members. These leaders realized teachers’ understanding of the purpose and design of the MBA as well as the standards assessed was essential to a successful implementation.

Grades 3 to 8 mathematics benchmark tests administered in spring 2010 to students in Summit districts were highly correlated with state Measures of Student Progress (MSP) tests. Pearson raw score to raw score correlations ranged from 0.709 to 0.802. High school scores were moderately correlated with High School Proficiency Exam (HSPE) scores at 0.358 (geometry) to 0.608 (algebra 2). All correlations were significant at the p = 0.01, Table 1b.

Background

In the absence of statewide development of mathematics interim (benchmark) test development, Office of Superintendent of Public Instruction (OSPI) staff in the District and School Improvement and Accountability Division (DSIA) developed multiple levels and forms of benchmark tests for grades K through high school in SY 2009-10 with the participation of mathematics teachers, ESD math specialists, and OSPI Teaching & Learning and Assessment Division partners. The team created three benchmark assessments per level (fall, winter, spring) from a common set of benchmark items and a model pacing framework. These tests matched the curriculum sequences of some participating districts. Project Director Shannon Edwards worked with Summit Project partner Teachscape to develop, deploy, score and analyze the tests administered fall, winter and spring within Summit districts (Edwards, 2010).
The Summit Project is a consortium of eight turnaround districts who worked in partnership with DSIA staff and four external contractors (Teachscape, BERC Group, CEE and WestEd). Despite development, administration, scoring and reporting issues, districts valued the information and persisted to benchmark 3. Math Technical Assistance Contractors with Special Expertise (TACSEs) and school-based Technical Assistance Contractors (TACs) provided many levels of support to turn the data into actionable information to shape instruction. The project continues in SY 2010-11.

Analysis

This study focused on the relationships between the math benchmark scores and grades 3 to 8 Measures of Student Progress and High School Proficiency Exam scores. The math benchmark tests were aligned with delivered units of instruction and administered following those units. Districts were permitted to shuffle test items to fit the local curriculum and to administer with alignment to local instructional calendars. It is not known if the several resulting test forms were equivalent in difficulty. TACSE Linda Elman performed item and test analyses (Elman, 2009). Districts which used the identical default benchmark forms were included in the current study. In contrast, state tests were administered within the same window for grades 3 to 8 and on the same days for high school.

The benchmark tests provided standards scores for teacher and school use. Test maps typically called for 1 to 8 items per standard. The tests were not intended to be summative and TACSEs counseled local districts to shun the use of total raw scores, and instead focus on standard level and distracter analysis data. Variability in standards scores is low, with as few as 1 to 3 raw score points for some standards.

This analysis used aggregate benchmark raw score test totals (Total) to relate to aggregate state test raw score and scaled score totals (Raw and Scale). Greater variability results by aggregating scores from all 21 to 30 items, providing more meaningful correlations.

Elman provided SPSS output from her 2009-10 math benchmark 1, 2 and 3 item analysis files. Records were identified by student ID numbers, largely unique within (but not across) districts. These were not OSPI-assigned SSIDs which were unique across districts. CEE Co-CEO Greg Lobdell provided MSPE and HSPE score files which included both SSID and District Student ID numbers in addition to many other demographic and subscore variables.

Results

Resulting correlations ranged from a low of \( r = 0.354 \) (high school geometry 1) to high of \( r = 0.803 \) (grade 7). The following tables display the relationships for each tested grade where there were both math benchmark and state tests scores. All correlations were significant at the \( p \geq 0.01 \). In some cases rows or columns do not add exactly due to exclusions such as students classified at level BA (Basic with IEP). Score ranges are noted in ( ) in the L1, 2, 3, 4 totals cells. Items were scored 1 (correct) or 0.

Discussion

Correlations were consistently in the \( r = 0.7 \) to 0.8 range in the MSP years, grades 3 to 8. Roughly half to two-thirds of the score variation in one test is explained by the variation in the other grades 3 to 8 tests as noted from the square of the ranges, \( r^2 = 0.50 \) to 0.64. The HS correlations were lower, ranging from \( r = 0.53 \) to 0.61 in Algebra 1 and 2 to \( r = 0.51 \) in Geometry. As little as \( r^2 = 0.125 \) of the variability in the geometry tests was explained. Students are routinely placed in HS classes according to their position in the traditional math sequence but the HSPE is only administered to the population at Grade 10. Tests for 11th and 12th graders are make-ups and are not reflected in the grade 10 results.
Table 2  
**Grade 3 spring 2010 math benchmark 3 correlation with MSP exam (maximum possible=30)**

<table>
<thead>
<tr>
<th>District</th>
<th>Benchmark Scores</th>
<th>Matched Scores</th>
<th>Correlation MSP Raw to Math Total</th>
<th>Correlation MSP Scale to Math Total</th>
<th>L1</th>
<th>L2</th>
<th>L3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mount Adams</td>
<td>93</td>
<td>69</td>
<td></td>
<td></td>
<td>21</td>
<td>27</td>
<td>16</td>
<td>5</td>
</tr>
<tr>
<td>Othello</td>
<td>291</td>
<td>290</td>
<td></td>
<td></td>
<td>75</td>
<td>63</td>
<td>93</td>
<td>55</td>
</tr>
<tr>
<td>Total</td>
<td>384</td>
<td>359</td>
<td>$r = 0.790$</td>
<td></td>
<td>96</td>
<td>90</td>
<td>109</td>
<td>60</td>
</tr>
</tbody>
</table>

$r^2 = 0.624$  
$r = 0.783$  
$r^2 = 0.612$  
$s = (5-21)$  
$s = (9-24)$  
$s = (11-28)$  
$s = (14-29)$  

Table 3  
**Grade 4 spring 2010 math benchmark 3 correlation with MSP exam (maximum possible=26)**

<table>
<thead>
<tr>
<th>District</th>
<th>Benchmark Scores</th>
<th>Matched Scores</th>
<th>Correlation MSP Raw to Math Total</th>
<th>Correlation MSP Scale to Math Total</th>
<th>L1</th>
<th>L2</th>
<th>L3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clover Park</td>
<td>953</td>
<td>939</td>
<td></td>
<td></td>
<td>301</td>
<td>225</td>
<td>292</td>
<td>117</td>
</tr>
<tr>
<td>Mount Adams</td>
<td>93</td>
<td>81</td>
<td></td>
<td></td>
<td>66</td>
<td>8</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Othello</td>
<td>271</td>
<td>271</td>
<td></td>
<td></td>
<td>96</td>
<td>68</td>
<td>62</td>
<td>45</td>
</tr>
<tr>
<td>Total</td>
<td>1317</td>
<td>1291</td>
<td>$r = 0.802$</td>
<td></td>
<td>463</td>
<td>301</td>
<td>360</td>
<td>163</td>
</tr>
</tbody>
</table>

$r^2 = 0.643$  
$r = 0.787$  
$r^2 = 0.619$  
$s = (4-25)$  
$s = (6-26)$  
$s = (12-24)$  
$s = (18-26)$  

Table 4  
**Grade 5 spring 2010 math benchmark 3 correlation with MSP exam (maximum possible=21)**

<table>
<thead>
<tr>
<th>District</th>
<th>Benchmark Scores</th>
<th>Matched Scores</th>
<th>Correlation MSP Raw to Math Total</th>
<th>Correlation MSP Scale to Math Total</th>
<th>L1</th>
<th>L2</th>
<th>L3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clover Park</td>
<td>963</td>
<td>862</td>
<td></td>
<td></td>
<td>261</td>
<td>225</td>
<td>244</td>
<td>129</td>
</tr>
<tr>
<td>Mount Adams</td>
<td>71</td>
<td>61</td>
<td></td>
<td></td>
<td>29</td>
<td>18</td>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>Othello</td>
<td>293</td>
<td>284</td>
<td></td>
<td></td>
<td>89</td>
<td>73</td>
<td>80</td>
<td>40</td>
</tr>
<tr>
<td>Total</td>
<td>1327</td>
<td>1207</td>
<td>$r = 0.709$</td>
<td></td>
<td>379</td>
<td>316</td>
<td>337</td>
<td>170</td>
</tr>
</tbody>
</table>

$r^2 = 0.508$  
$r = 0.707$  
$r^2 = 0.500$  
$s = (3-19)$  
$s = (6-19)$  
$s = (6-21)$  
$s = (12-21)$  

Table 5  
**Grade 6 spring 2010 math benchmark 3 correlation with MSP exam (maximum possible=26)**

<table>
<thead>
<tr>
<th>District</th>
<th>Benchmark Scores</th>
<th>Matched Scores</th>
<th>Correlation MSP Raw to Math Total</th>
<th>Correlation MSP Scale to Math Total</th>
<th>L1</th>
<th>L2</th>
<th>L3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mount Adams</td>
<td>74</td>
<td>72</td>
<td></td>
<td></td>
<td>56</td>
<td>12</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Othello</td>
<td>264</td>
<td>74</td>
<td></td>
<td></td>
<td>98</td>
<td>70</td>
<td>72</td>
<td>23</td>
</tr>
<tr>
<td>Sunnyside</td>
<td>602</td>
<td>426</td>
<td></td>
<td></td>
<td>163</td>
<td>123</td>
<td>92</td>
<td>44</td>
</tr>
<tr>
<td>Total</td>
<td>940</td>
<td>736</td>
<td>$r = 0.772$</td>
<td></td>
<td>317</td>
<td>205</td>
<td>168</td>
<td>67</td>
</tr>
</tbody>
</table>

$r^2 = 0.596$  
$r = 0.757$  
$r^2 = 0.573$  
$s = (2-21)$  
$s = (2-18)$  
$s = (4-23)$  
$s = (71-24)$  

Continued on next page
Table 6
Grade 7 spring 2010 math benchmark 3 correlation with MSP exam (maximum possible=23)

<table>
<thead>
<tr>
<th>District</th>
<th>Benchmark Scores</th>
<th>Matched Scores</th>
<th>Correlation MSP Raw to Math Total</th>
<th>Correlation MSP Scale to Math Total</th>
<th>L1</th>
<th>L2</th>
<th>L3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mount Adams</td>
<td>69</td>
<td>67</td>
<td>34</td>
<td>18</td>
<td>13</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Othello</td>
<td>257</td>
<td>256</td>
<td>65</td>
<td>42</td>
<td>91</td>
<td>55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sunnyside</td>
<td>310</td>
<td>308</td>
<td>82</td>
<td>71</td>
<td>108</td>
<td>47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>636</td>
<td>632</td>
<td>r =0.803</td>
<td>r^2 =0.645</td>
<td>181</td>
<td>131</td>
<td>212</td>
<td>104</td>
</tr>
</tbody>
</table>

Table 7
Grade 8 spring 2010 math benchmark 3 correlation with MSP exam (maximum possible=29)

<table>
<thead>
<tr>
<th>District</th>
<th>Benchmark Scores</th>
<th>Matched Scores</th>
<th>Correlation MSP Raw to Math Total</th>
<th>Correlation MSP Scale to Math Total</th>
<th>L1</th>
<th>L2</th>
<th>L3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mount Adams</td>
<td>61</td>
<td>59</td>
<td>23</td>
<td>11</td>
<td>18</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Othello</td>
<td>166</td>
<td>165</td>
<td>28</td>
<td>36</td>
<td>51</td>
<td>48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sunnyside</td>
<td>392</td>
<td>389</td>
<td>171</td>
<td>71</td>
<td>101</td>
<td>45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>619</td>
<td>613</td>
<td>r =0.721</td>
<td>r^2 =0.520</td>
<td>222</td>
<td>118</td>
<td>170</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 8
Grade HS (10) spring math benchmark 3 correlation with HSP Exam (maximum possible=23)

<table>
<thead>
<tr>
<th>District</th>
<th>Benchmark Scores</th>
<th>Matched Scores</th>
<th>Correlation MSPE Raw to Math Total</th>
<th>Correlation MSPE Scale to Math Total</th>
<th>L1</th>
<th>L2</th>
<th>L3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mount Adams</td>
<td>121</td>
<td>75</td>
<td>r =0.526</td>
<td>r^2 =0.276</td>
<td>149</td>
<td>374</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>Othello</td>
<td>420</td>
<td>199</td>
<td>r =0.608</td>
<td>r^2 =0.370</td>
<td>55</td>
<td>58</td>
<td>60</td>
<td>36</td>
</tr>
<tr>
<td>Sunnyside</td>
<td>759</td>
<td>327</td>
<td>r =0.358</td>
<td>r^2 =0.128</td>
<td>80</td>
<td>44</td>
<td>18</td>
<td>1</td>
</tr>
<tr>
<td>Algebra 1</td>
<td>223</td>
<td>194</td>
<td>r =0.522</td>
<td>r^2 =0.272</td>
<td>149</td>
<td>374</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>Algebra 2</td>
<td>244</td>
<td>210</td>
<td>r =0.615</td>
<td>r^2 =0.378</td>
<td>55</td>
<td>58</td>
<td>60</td>
<td>36</td>
</tr>
<tr>
<td>Geometry 1</td>
<td>154</td>
<td>143</td>
<td>r =0.354</td>
<td>r^2 =0.125</td>
<td>80</td>
<td>44</td>
<td>18</td>
<td>1</td>
</tr>
<tr>
<td>All 3 Courses</td>
<td>621</td>
<td>547</td>
<td>r =0.511</td>
<td>r^2 =0.261</td>
<td>284</td>
<td>135</td>
<td>88</td>
<td>39</td>
</tr>
</tbody>
</table>

Continued on next page
The tests were designed for different purposes. Benchmark tests are classroom instruments, mainly for teacher use with students and school/district use at the program level. State tests were designed primarily for accountability at the school and district level, not for classroom use as the preliminary scores are not available until school is out (excepting some high school scores). The data from state tests may guide students towards various interventions or recognition, but it is not valuable in the weeks of instruction immediately ahead. The underlying curricular targets, what will become the Common Core Standards, are common across schools and districts. That is not to say the delivered curriculum is common. Any fair observer would note significant differences in instruction within and between schools and districts.

What are "strong" correlations between tests measuring similar constructs and targets?

Washington moved to the Washington Assessment of Student Learning (WASL) over 10 years ago and correlations between WASL math and the Iowa Test of Basic Skills, the prior state test, were said to be fairly high \((r=0.77)\) showing a substantial overlap (Joireman & Abbott, 2001). A more recent analysis of the relation of WASL scores to college placement tests in play in Washington public colleges stated that "commonly accepted values" for correlations were small at \(r=0.1\), medium at \(r=0.3\) and large at \(r=0.5\) and higher (McGhee, D., 2003). With \(n's\) greater than 100 for all comparisons, there is little worry about normality assumptions in the data (StatSoft, 2010).

All the correlations are at least modest and the MSP correlations would be judged high by most analysts. When comparing tests, there is concern that correlations which are too high, above 0.95, for example, may indicate that the overlap is so great between tests than one (or the other) is not necessary. Benchmark tests serve a different purpose than the state tests. An argument can be made that the high correlations demonstrate that similar concepts or learning targets are being tested. Similarly, the strong correlations help make the case that the benchmark tests are valid as they produce scores that vary like those in other math tests.

However, we have not examined which students were excluded from benchmark testing or simply withdrew between one session and the other. Unlike state summative tests, math benchmark testing is not held to the same levels of secure document handling and standardized administration. It is not known if teachers prepared or assisted students differently than in state test administration. And there were considerable challenges in scoring the tests, particularly in the fall and winter sessions.

Implications

- **Could benchmark scores be used to predict performance on MSPE of HSP tests?** Predicting outcomes is a probabilistic enterprise--there is no absolute certainty that a score on one test will always predict a score on the other test. However, higher raw scores on the benchmark tests increase the certainty that students will score higher on state tests. If predications are made at the student level, they should be accompanied by a statement about the certainty of the prediction. For example, "Students with scores of XX to XX are likely to meet standard XX times out of 100." We have not performed this analysis.

- **Does a fall score tell me more about what the student has just learned or about what they are going to learn?** Predicting remains probabilistic. It is not clear how closely aligned the tests are with the delivered instruction in each classroom. However, if the new items are scaled and given a known difficulty rating, growth could be measured from one benchmark score to the next. The vendor was not contracted to perform this work.

- **Could teachers use the benchmark standard scores to predict performance on the state test standard scores?** When correlating a restricted score range (a few benchmark items) to another restricted score range (a few state test items), correlations tend to grow weaker. Those analyses were not performed.

- **Might the HS end of course tests in 2011 correlate more strongly to the new benchmark tests?** We remain curious to see if, for example, student scores from an Algebra 1 end of course test relate strongly to the Algebra items in the benchmark test near the close of semester 2. We plan to perform this analysis.
References


-Peter Hendrickson is a Management TACSE with OSPI and edits the Journal.

-Shannon Edwards directs the mathematics improvement efforts for the Secondary Education and School Improvement division at OSPI. Contact her at shannon.edwards@ospi.k12.wa.us.
If you’ve ever watched Hans Rosling take four minutes to share data on the life expectancy and wealth of 200 countries over 200 years (BBC, 2010), you understand the power of data visualization. Rosling uses the Trendalyzer software developed by his Gapminder Foundation to animate and communicate a large amount of data. In 2007, Google acquired the Trendalyzer software (Gapminder, n.d.) and now makes a version of that software available to you. Google's version, Motion Chart, is available in several forms, including as a gadget on a Google Site, a gadget on your own website, or as code embedded on your own website. I used my district's WASL/MSP/HSPE data on a Google Site: https://sites.google.com/site/mukilteodata/. Using a gadget in a Google Docs spreadsheet (the easiest form to implement), it is possible to communicate large amounts of data involving multiple variables over time. Here’s how you can quickly utilize a Google Motion Chart in your Google Docs spreadsheet.

![Figure 1: Data organized using Microsoft Excel](image)

**Organize Your Data with Excel**

First, select a data source that includes a measurement of time, such as standardized test data. Next, organize the data so that the first column contains the variable that you want to track. In my case, it was the names of the schools in my district. It’s important to be consistent in your titles; the gadget tracks the data using those titles and displays it next to the bubbles on your chart. The second column must include dates, or another measure of time. Be sure to format the cells as numbers (not numbers stored as text). *(Hint: If you see that pesky green triangle in the corner of the cell--those are numbers stored as text. You can quickly convert the cell(s) to a proper number by highlighting your list of misbehaving numbers, clicking the exclamation mark, and selecting ‘Convert to Number’ from the dropdown.)* Subsequent columns should contain other data to track. Use clear column headings (labels), as these will be the variables you select in your finished motion graph. I included test scores, grade levels and demographic information.

**Upload Data to Google Docs**

Create a Google Docs Account if you don’t already have one (www.docs.google.com). Log into Google Docs and upload your Excel document. As you select your document, be sure to select the check box next to, “Convert documents, . . . to the corresponding Google Docs formats.”

**Insert the Gadget**

Open your spreadsheet in Google Docs and add a sheet, using the + button in the light blue band at the bottom of your browser window. From the Insert menu, select “Gadget...” then scroll to find Motion Chart and click “Add to Spreadsheet.” When the Gadget Settings box appears, select the button to the right of Range, select the sheet with your data, highlight your data and click OK. Enter a title in the Title box, but leave the Default state box blank for now. Click “Apply & close” and navigate to the sheet with your motion chart.

Continued on next page
Adjust the Settings on Your Gadget

At this point, your motion chart is functional. You may want to enlarge the chart by selecting and dragging its lower right corner and experiment with the format and variables to display. Once you have settled on the format and variables to display in the chart’s default state, click the wrench in the lower right corner. When the dialog box appears, click “Advanced” at the bottom of the box, and then the “Advanced” that pops up right above that. Highlight and copy the text in the State string box then close the settings box. In the upper right corner of the gadget, click on the title of your chart and select “Edit Gadget.” Paste the state string into the Default State box in the Gadget Settings and click “Apply & close” to establish the default state of your gadget. Now, whenever your page loads or refreshes, the motion chart will default to the settings you have selected.
More Motion Chart Possibilities

The simplest way to implement a motion chart is as a gadget within a Google Docs spreadsheet. If you want a little more polished look, you can create a Google Site and add a gadget to it. Start by uploading your data to Google Docs, open the spreadsheet and copy its URL. Next, while in edit mode of your Google Sites webpage, select “More Gadgets” from the Insert menu and search for Motion Chart. Paste the URL into the “data source url” box to start setting up your motion chart. Similarly, you can add a gadget to your own site (http://www.google.com/ig/directory?type=gadgets&url=www.google.com/ig/modules/motionchart.xml) or, if you are more adventurous, you can embed the motion chart right in your webpage’s code (http://code.google.com/apis/visualization/documentation/gallery/motionchart.html).

Discussion

While the Google motion chart may not be helpful for all types of data, it can be useful for quickly communicating trends in a large amount of data with multiple variables and a time component. It’s also an opportunity for others to interact with the data and modify the chart to show the data and relationships that interest them.

References


--Bruce Denton is Title 1 Facilitator in Mukilteo School District. Contact him at DentonBJ@mukilteo.wednet.edu.
Techniques for Effective Use of Color in Data Display
By Andrea Meld, Ph.D.

Graphics expert Edward Tufte devotes an entire chapter to the use of color and information in his often cited book, Envisioning Information (1990). He explains that “human eyes are exquisitely sensitive to color variations (p. 81).” The Yale emeritus professor points out that with training, it’s possible for color experts to distinguish among 1,000,000 colors, and when tested, many people can distinguish among as many as 20,000 colors while looking at adjacent color chips. Yet when it comes to the use of color for encoding data or abstract information, using too many colors or the wrong colors can produce poor or even disastrous results. He adds, “The often scant benefits derived from coloring data indicate that even putting a good color in a good place is a complex matter. Indeed, so difficult and subtle that avoiding catastrophe becomes the first principle in bringing color to information: Above all, do no harm (p. 81).”

A Brief History of Color Theory and Representation

The history of color theory as a subject of study starts with Aristotle, and includes the writings of Leonardo Da Vinci, Goethe, and Newton, who each developed models to represent the properties of color. Sir James Clerk Maxwell, a Scottish physicist, developed a triangular chart based on his research on the electromagnetic properties of light, with red, green, and blue as primary colors in each corner in 1872 (see Maxwell Model.gif).

Albert Munsell, an American art instructor, developed a three-dimensional, non-symmetrical model of color space that remains of influence to artists, manufacturers of paint, and others who mix pigments to produce a wide range of colors. The Munsell model takes into account that pure hues (red, yellow, green, blue, for example) vary in degrees of lightness and darkness (see Munsell Model.jpg).

As color modeling became more important with the advent of color photography and film, the Commission Internationale de E’clairage (CIE) attempted to set up an international standard for the measurement of color in 1931, based on Maxell's triangle, choosing a particular red, green, and blue from which to generate all the colors (see CIE Chromaticity Diagram.gif). The result was the CIE Chromaticity Chart, and a newer version is used to measure and quantify the light produced by computer phosphor guns. (For a more detailed on the history of color modeling, see ElectronicColor.html.)

Regardless of color model, any and all colors can be represented in terms of three variables: either by hue, saturation and value, or by amounts of red, green, and blue. These three variables are also commonly encoded for web displays using hexadecimal notation, a base 16 number system in which 0 – 9 represent values of zero to nine and letters A – F represent values of 10 to 15. (See Table 1.)

Table 1
Saturated colors with corresponding hexadecimal and decimal codes

<table>
<thead>
<tr>
<th>color name</th>
<th>Hex code</th>
<th>Decimal code</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R  G  B</td>
<td>R  G  B</td>
</tr>
<tr>
<td>Crimson</td>
<td>DC 14 3C</td>
<td>220 20 60</td>
</tr>
<tr>
<td>Gold</td>
<td>FF D7 00</td>
<td>255 215</td>
</tr>
<tr>
<td>Yellow</td>
<td>FF FF 00</td>
<td>255 255</td>
</tr>
<tr>
<td>Lime</td>
<td>00 FF 00</td>
<td>0 255</td>
</tr>
<tr>
<td>Green</td>
<td>00 80 00</td>
<td>0 128</td>
</tr>
<tr>
<td>Blue</td>
<td>00 00 FF</td>
<td>0 0</td>
</tr>
</tbody>
</table>

Color monitors use three different types of phosphors that emit red, green, or blue light when activated. Various combinations of these phosphors in different intensities produce a multitude of different colors. Red, green and blue are the primary colors; other colors are produced by combining different intensities of these primary colors. For example, yellow is produced by a combination of red and green light. A scale of 0 to 255 for red, green and blue specifies a particular color. Your computer screen is initially black. As red, green, and blue lights are added, the screen becomes brighter and lighter. Figure 1 displays an array in Excel. When there is 100% intensity of the red, green, and blue phosphors the screen is white.

Figure 1
Color specification in Excel either by cursor or red, green, and blue scales

Continued on next page
Suggestions for Using Color in Data Display

Use White for Background Color

Colors used for data display are generally designed to be printed on white paper; thus, in digital display a white background is preferable, and will also make it easier to select colors that will work best both in electronic and print form. There are also perceptual advantages to using white as a background. Color perception in humans sees hues (blue or red, for example) and shades of the same hue (light blue, dark blue) in relation to white. We adjust our focus to some degree to different colors. A white background provides a greater stability to focus on (Stone, 2006). Peltier (2008) urges Excel users to change the default “ugly gray plot area” to white in Excel charts, but recommends using light gray for lines, borders, and other supporting chart elements.

In the past, slide projectors and early models of digital projectors produced relatively dim color displays and required a dark room for viewing. In a darkened setting, a light text on a dark background is easier to see, especially with vision adapted to the dark. However, digital projections produced by modern equipment should be bright enough to view in regular daylight. For practical purposes, the only reason to use light text on a dark or black background is if the viewer is seeing a display in the dark (Stone, 2006).

Use Color Carefully and with Restraint

• Use color to communicate rather than decorate your data.

• Use a range of colors within a single hue, from pastel to dark, to show order or sequence from low to high. (Note: see section on color blindness for caveats.)

• Use soft, natural colors to display most of the information, and bright or dark colors to call attention to your data. These colors are also easier to duplicate in print, and allow you to use brighter, more saturated tones for highlighting. (See Appendix for additional color samples and RGB codes.)

• Although colors of medium shade work best for bars or larger areas, they do not show up as well for small data points or lines.

• For small data points or lines, use bright or dark colors, or enlarge the lines or points so that they are easier to see.

• Avoid using a combination of reds and greens in the same display so that most people who are colorblind can see the color-coding. (See Figure 2.) You will also avoid distracting visual distortions for those with normal color vision.

Figure 2
*Saturated red and green in the same display may create “shimmer” or other visual illusions.*
• Avoid using distracting visual effects in graphs, as shown in Figure 3.

Figure 3
*Unnecessary and distracting visual effects in a column graph.*

• Use a single set of ordered colors in the same hue to represent equal intervals of a variable, or intervals that are perceptually equivalent. For example, population density, income distribution, number of children of school age, etc.

• Use a dual-ordered palette to show variables that may be positive or negative, with the lightest colors representing the middle or zero point (Figure 4). Dual-ordered palettes can also be used to show demographic characteristics such as percent voting for different political parties.

Figure 4
*Ordered Colors in a Single Hue (left) and examples of Dual-Ordered Palettes (right)*

*Continued on next page*
According to Arditi (2011), many forms of color blindness are genetic and more prevalent in men (about 8%) than in women (about 0.4% to 2%). This means that if you work in a building with 400 people, for example, you may have 33 or so coworkers who have difficulty interpreting color. Color blindness can interfere with color-related tasks in certain occupations, including the military. In addition, the cornea tends to become more yellow with age, reducing the ability to see and distinguish cool colors, especially those in the blue to purple range. This range of colors may appear instead as green or brown tones. Even in their middle years, “most people over the age of 45 will experience some kind of vision loss that makes distinguishing among different hues (red and green, for example) and different shades of the same hue (light green and dark green) more difficult” (Arditi, 2011). This has implications for the use of color and color contrasts in data display, as will be discussed further.

Color blindness has several forms, most common is the lack of sensitivity to red or green, although some people can perceive red and green but lack sensitivity to blue. See Figure 5.

![Figure 5](image)

Top row: colors as perceived with normal vision; Middle row: appearance if lacking red or green sensitivity; Bottom row: appearance if lacking blue sensitivity

Aires Arditi, a vision researcher at Lighthouse International, suggests some basic rules for color selection and color combinations so that data displays work better for those with color blindness or vision loss:

1. Foreground and background colors should differ in lightness value, and it is generally preferable to use dark text on light background, rather than the other way around.

2. When using colors that are adjacent on the color wheel, for example, red-orange-yellow, or purple-blue-green, colors should be of different lightness levels, even if they differ in terms of hue and saturation.

3. People with color vision deficits tend to perceive less contrast than those with normal vision. By lightening light colors and darkening dark colors in your display, visual accessibility will increase.

4. Avoid using colors that are across from each other on the color wheel, such as red-green, yellow-purple, and orange-blue, in combinations where one color is directly on the other color.

Another overall guideline for accessibility is to avoid the use of color alone to convey information. Instead, provide redundant means of conveying information. As Stone (2006) recommends, “get it right in black and white.” Ideally, all-important information should be legible in your display, even if reproduced in shades of grey (Stone, 2006).
References and Resources


Continued on next page

### Appendix: Sample Colors for Data Display with Corresponding RGB Codes

<table>
<thead>
<tr>
<th>Color</th>
<th>Red</th>
<th>Green</th>
<th>Blue</th>
<th>Color</th>
<th>Red</th>
<th>Green</th>
<th>Blue</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>180</td>
<td>250</td>
<td>250</td>
<td></td>
<td>250</td>
<td>150</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>200</td>
<td>250</td>
<td></td>
<td>215</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>126</td>
<td>134</td>
<td>242</td>
<td></td>
<td>145</td>
<td>90</td>
<td>145</td>
</tr>
<tr>
<td></td>
<td>58</td>
<td>160</td>
<td>230</td>
<td></td>
<td>90</td>
<td>155</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>150</td>
<td>215</td>
<td>200</td>
<td></td>
<td>250</td>
<td>231</td>
<td>114</td>
</tr>
<tr>
<td></td>
<td>165</td>
<td>200</td>
<td>125</td>
<td></td>
<td>228</td>
<td>244</td>
<td>165</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>200</td>
<td>120</td>
<td></td>
<td>168</td>
<td>168</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>224</td>
<td>160</td>
<td>114</td>
<td></td>
<td>200</td>
<td>120</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>230</td>
<td>130</td>
<td>95</td>
<td></td>
<td>185</td>
<td>120</td>
<td>138</td>
</tr>
<tr>
<td></td>
<td>220</td>
<td>160</td>
<td>60</td>
<td></td>
<td>158</td>
<td>72</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>10</td>
<td></td>
<td>240</td>
<td>150</td>
<td>150</td>
</tr>
</tbody>
</table>

--Andrea Meld is a data analyst with OSPI and edits WERA’s The Standard Deviation newsletter. Contact her at andrea.meld@k12.wa.us.

Manuscript received March 8, 2011  
Final revision received April 20, 2011  
Accepted April 20, 2011
Dr. D’ Lema: A WERA Journal Forum on Data and Research Ethics

In a complex world of data and information, misinformation abounds. This column is intended as a meeting place for discussion about data and research ethics, and advocacy.

The July 2010 Dr. D’Lema column reviewed the NCES Code of Data Ethics, which are summarized below:

**Core Principles of the Code of Data Ethics**

**The Integrity Canon**
1. Demonstrate honesty, integrity, and professionalism at all times.
2. Appreciate that, while data may represent attributes of real people, they do not describe the whole person.
3. Be aware of applicable statutes, regulations, practices, and ethical standards governing data collection and reporting.
4. Report information accurately and without bias.
5. Be accountable and hold others accountable for ethical use of data.

**The Data Quality Canon**
1. Promote data quality by adhering to best practices and operating standards.
2. Provide all relevant data, definitions, and documentation to promote comprehensive understanding and accurate analysis when releasing information.

**The Security Canon**
1. Treat data systems as valuable organizational assets.
2. Safeguard sensitive data to guarantee privacy and confidentiality.

In this installment of Dr. D’Lema, the Security Canon, and in particular, the obligation to safeguard sensitive data and to insure the privacy and confidentiality of student records is explored. In addition, specific questions about data privacy can be obtained at a new website, the Privacy Technical Assistance Center, [http://nces.ed.gov/programs/Ptac/Home.aspx](http://nces.ed.gov/programs/Ptac/Home.aspx) a service of NCES, which also provides information about best practices concerning privacy, confidentiality, and security as we embark on this age of longitudinal data systems.

Other data ethics canons will be considered in future columns, as well as new standards for educational and psychological testing (AERA, APA and NCME), a revision of the 1999 standards. Draft versions of the new standard and other materials are available at [http://teststandards.org](http://teststandards.org).
Data Privacy: What it is and Why it matters
By Andrea Meld, Ph.D.

The Code of Data Ethics, published by the National Forum on Educational Statistics encompasses nine cannons (Purwin, T., McMurtrey, C., Metcalf, S., Petro, J., Rabbitt, L., & Uhlig, D., February, 2010). This article explores the last cannon, concerning data security, “Safeguard sensitive data to guarantee privacy.” Some of us are old enough to remember the anxiety of seeing your exam or course grades posted on the office door of teachers or professors, with your last name in one column and the decimal or letter grade in another, in full view of other students, teachers and anyone else who happened to walk by. This type of grade posting, once commonplace, is strictly prohibited by the Family Educational Rights and Privacy (FERPA), enacted by Congress is 1974. Today, as teachers, administrators, and data professionals we may be responsible for the data privacy and security of a classroom of students, or even millions of electronic student records.

Three Ethical Injunctions

In his AERA presentation on the ethical basis of human research protections and federal regulations, Ivor Pritchard evoked three ethical injunctions that also apply directly to data ethics and data privacy:

• The Principle of Respect for Persons, “Showing Respect”
• The Principle of Beneficence, “Do Good.”
• The Principle of Justice, “Be Fair.”

Protecting the Privacy of Student Records: Guidelines for Education Agencies (Cheung, Clements, and Pechman, July, 1997), summarizes principles for those who collect, use, or provide personal information, which correspond to these injunctions:

Respect

“Personal information should be acquired, disclosed, and used only in ways that respect an individual’s privacy (p. 8).” It should not be misused, or inappropriately changed or destroyed. When deciding whether to collect or disclose personal information, assess the potential impact on personal privacy.

When education agencies collect data from people, they should tell them “1) why the information is being collected; 2) what the information is expected to be used for; 3) what steps will be taken to protect its confidentiality, integrity, and quality; 4) the consequences of providing or withholding information; and 5) any rights of redress (p. 8).”

Do Good (and by implication, do no harm)

Education agencies should collect and maintain only the types of information that are intended and expected to support current or planned programs and services. Technical procedures need to be in place to protect the confidentiality and integrity of personal information.

Fairness

Personal information should only be used in ways that match the individual’s understanding. Individuals have the right to correct personal information that is incorrect and the right to redress if personal information is used improperly or causes them harm.

These basic ethical injunctions should take us far in making sure that procedures are respectful of students and student records, that we establish and follow policies that benefit students and their families, and that we support social rights and social justice.

A Very Brief History of Past and Present of Privacy Rights

The concept of privacy, the right to be free from physical intrusion and to share information about oneself as one chooses, may differ across time and cultures. Some languages lack a word to denote privacy, and some argue the concept of privacy can be translated into Russian language only with great difficulty (Anderman & Rogers, 2003).

Privacy as a legal right has origins in British common law, and was codified into American Law by the Fourth Amendment of the Bill of Rights,

www.loc.gov/rr/program/bib/ourdocs/billofrights.html, which prohibits “unreasonable searches and seizures.” In 1974, roughly 200 years later, Congress passed FERPA, which guarantees the rights of parents and students to confidentiality and fairness with regard to the maintenance and use of student records. The provisions of FERPA, an important landmark, must be strictly followed by public schools, school districts, colleges, and other educational institutions that receive federal funding. “The school district is responsible for ensuring that all parents and eligible students are afforded all the rights provided them by FERPA (2006).”

The Future of Data Privacy

No Child Left Behind brought about a vast expansion in the quantity of information about student, school, and school districts that are reported to parents and the general public, with student outcomes reported for reading and math scores, at multiple grade levels, as well as by the categories of race/ethnicity, and

Continued on next page
participation in programs aimed at students with special needs, limited English proficiency, and low income. In addition, states frequently report data on attendance, course taking, and graduation and dropout rates. This type of information is generally reported in aggregate, at the school, district, group, or state level. “These reports offer the challenge of meeting the reporting requirements while also meeting legal requirements to protect each student’s personally identifiable information (Seastrom, November, 2010, p. 1).

The current interest in models of student growth, value-added teacher evaluations, and other types of research following individual trends has prompted an increase in development of student longitudinal data bases, including those at the state level. With newer models have come new concerns and technical proposals for safeguarding student privacy. NCES has published a series of technical briefs addressing these issues and providing guidance for statewide longitudinal data systems on protecting privacy and confidentiality (Seastrom, November 2010, p. 1) data stewardship (Seastrom, November 2010, p. 2) and statistical methods for protecting personally identifiable information in electronic student records (Seastrom, December, 2010, p. 3).

References and Resources


--Andrea Meld, Ph.D. is a data analyst at OPSI and editor of The Standard Deviation.
Data Privacy: Selected Key Terms and Definitions

**Privacy.** A personal right to be free from unwanted intrusion. Privacy rights of students and parents are violated when personally identifiable information is disclosed to others without appropriate consent, or when they are asked for information by those who have no legal basis to do so.

**Confidentiality.** The obligation of teachers and other professionals to refrain from sharing confidential information about students and their families to unauthorized parties. Confidential information includes private, sensitive, and personally identifiable information.

**Private Information.** Data considered very personal and not for public release, nor accessible without an established "need to know," for example, course history, grades, test scores, medical information, unexcused absence and disciplinary action.

**Sensitive Information.** Information about an individual student that might have a negative effect on that student if improperly disclosed.

**Personally Identifiable Information.** Information that can be used to expose a student's personal identity, or that can be used in combination with other information (e.g., by linking records) to identify a student, such as name, address, identification numbers, or any other information that could be used to identity a student.

**Directory Information.** Information in a student’s record that generally would not be considered harmful or an invasion of privacy if disclosed.

**Security.** Policies and procedures that ensure the confidentiality and integrity of student records, and that allow only authorized disclosure of confidential student data.
WERA Book Reviews

This issue’s reviews are an eclectic but interesting collection of reflections on a trio of books that should appeal to various interests.

**Fro Mesendick and Scott Taylor** share their insights on a new co-release from the Council of Chief State School Officers and the Association of Test Publishers, *Operational Best Practices for Statewide Large-Scale Assessment Programs*.

**Jack Monpas-Huber** offers his thoughtful point of view on Edward Tufte’s most recent exploration of data displays, *Beautiful Evidence*.

**Heather Rader** rounds off the trio with some personal reflections on John Medina’s very popular *Brain Rules: 12 Principles for Surviving and Thriving at Work, Home and School*. Medina presented on the same topic at this Spring’s WERA conference.

-Phil Dommes, Book Review Editor
Operational Best Practices for Statewide Large Scale Assessment Programs Authored by Council of Chief State Offices & Association of Test Publishers
Reviewed by Frosyne Mensendick and J. Scott Taylor

Operational Best Practices for Statewide Large-Scale Assessment Programs is the product of a working group assembled by the Association of Test Publishers (ATP) and sponsored by the Council of Chief State School Officers (CCSSO). The working group was charged with the task of developing a set of recommended best practices that could be used to enhance state assessment programs conducted under the No Child Left Behind Act (NCLB), and was comprised of both assessment industry and state assessment office representatives. Initial discussions in regard to this effort took place in 2006, with the final version of Best Practices published in 2010.

This review attempts to view Best Practices from these perspectives:

- client agency: state assessment office which contracts with assessment industry companies to administer an NCLB-compliant large-scale assessment
- service provider: assessment publishing companies
- educator: teachers, school counselors, test directors and others in local education agencies (LEAs) charged with administration of large-scale assessments and with using assessment results to inform local instruction and policy

Best Practices outlines assessment development and contract management practices already in regular use for a number of years by most major assessment publishing companies, those practices having evolved through decades of practical experience and periodic changes in assessment requirements, design, and delivery. Since the late 1990s and prior to NCLB, these best practices have become fairly standardized due to the industry’s emphasis on obtaining third-party project and program management certification, such as the Project Management Institute’s Project Management Professional (PMP) and Program Management Professional (PgMP), for program and project managers directly assigned to large-scale assessment contracts such as those related to NCLB. Some companies have gone a step further by requiring management certifications for functional group managers (e.g., printing, distribution, scoring, reporting, and information technology). NCLB also influenced the industry to adjust and update best practices in light of new realities faced by their client state agencies.

On the other hand, prior to NCLB, a client agency already well-versed in sound and efficient assessment development practices was an exception. NCLB, due to its inherent accountability and legal challenges, was a catalyst for client agencies to improve and further standardize their own assessment development practices. A number of state agencies now employ program managers to directly manage the agencies’ work on assessment development and administration, and input from those managers is well-represented on the ATP working group. One chapter, for example, is dedicated entirely to the client-centric task of assessment program procurement.

Developed by client agency and industry representatives, Best Practices primarily facilitates high-level management (state agency and publisher) of assessment development and implementation and offers little that would be useful to teachers and administrators in the daily work of administering student assessments. While offering best practices to client agencies and assessment publishers in regard to development of interpretive guides, information about using assessment results to determine effectiveness of instruction is not included. However, because school counselors, administrators, and teachers are often called upon by parents and the press to describe aspects of NCLB assessments that will be or have been administered to their students, Best Practices, even in its current form, would be an excellent source of useful information to inform that discussion.

Six of the 21 chapters in Best Practices address development and administration of traditional paper-pencil assessments, including form development, booklet construction and manufacturing, packaging and transportation, and post assessment retrieval and staging of scannable response media. Outlined in these chapters are comprehensive, time-proven best practices related to assessment administration using both physical question and response media.

Chapter 15 addresses online assessment, and with more and more states and companies employing online assessments, a need for best practices related to computer-based assessment is becoming increasingly important. Still, computer-based large-scale assessment is relatively new in terms of implementation and what is presented in this chapter are “… current practices as opposed to defined state-of-the-art practices.” The outline of suggested best practices in Chapter 15 is primarily centered on technical issues (e.g., infrastructure, usability, capabilities) and customer support (e.g., training, technical assistance) and is as well-considered and thorough as it can be given the relative newness of computer-based assessment. A note
in the introduction to Best Practices indicates that the ATP will recruit a new working group in 2011 to collect and develop new ideas and considerations (for the complete book). We learned as we were completing this review that a group is meeting to review and consider edits and additions to the book. It should be safe to assume that one of the outcomes will be an expanded set of best practices for online assessment that will have had the benefit of much more real world experience. Aside from updating and expanding this particular chapter, other chapters could be updated as well to include information about online assessment. For example, information about distribution of test forms that is now focused on physical materials could be expanded to suggest best practices for deployment of online test forms, including form sampling. We look forward to seeing how treatment of innovative item formats will be reflected through best practices.

Some highlights:

• Chapters 2 (item development) and 4 (test form construction) effectively address two aspects of work that, when planned or managed poorly, will almost always result in schedule delays and unexpected/unplanned costs, and present an opportunity for social and even legal discord between a client agency and its service provider. For example, a client agency may employ multiple reviewers of items and forms, but without effective scheduling and oversight of the review team’s work, it is entirely possible—perhaps even likely—that the client agency will receive multiple versions of item and form edits to the service provider, resulting in confusion and/or unplanned, additional review rounds. On the service provider side, managers have sometimes been guilty of building, with little or no contingency planning, assessment development and implementation schedules based on mistaken, best-case assumptions about the client agency’s capabilities. The best practices presented in these two chapters, if adopted in advance (and as applicable) by both sides, should be most effective in ensuring a successful assessment implementation

• In the days of item cards and file cabinets, item banking was not much of an issue. However, as electronic data systems and software design began to proliferate and evolve, online item bank development and maintenance became a fairly regular source of friction between client agencies and service providers. This friction was primarily a product of some service-provider proposals that described yet-undeveloped or only partially developed “Cadillac” item bank capabilities combined with a client tendency toward unrealistic expectations. For contracts requiring development and maintenance of online item banking, Chapter 3 should be particularly helpful in centering the expectations of both the client agency and the service provider.

For those client agencies and service providers already well-versed in program management, there is little new ground here; the best practices described could well have been excerpted from well-done client agencies’ Requests for Proposals (RFPs) and service providers’ proposals. However, for client agencies and service providers that recognize a need for stronger program management or that wish to avoid complacency, Best Practices represents its working group’s recommendations for development, implementation, and maintenance of strong, NCLB-compliant assessment programs from both the client and service provider perspectives.

Best Practices is, at its best:

• an instructional piece, offering in each chapter (after a brief introduction) the working group’s conclusions as to best practices in outline form, with bullet lists included in many of the outline elements. Each element of the outline is more than adequately informative, yet concise and brief enough that the outline as a whole can be grasped and followed easily.

• an excellent learning tool.

• a convenient and well-organized source of reference for post-award dialogue between the client agency and its contracted service provider(s).

• complete enough in that it could well be referenced in future RFPs and proposals as being the definitive guide for post-award assessment implementation.


--Fro Mesendick and J. Scott Taylor were both employees of Pearson working to support state assessment contracts but each has held the role of educator/consumer of services in previous lives.

Fro is a WERA past president. Scott recently retired. Contact them at froyne.mensendick@pearson.com or aquavelvis@sbcglobal.net
Beautiful Evidence by Edward Tufte
Reviewed by Jack Monpas-Huber

Surely anyone who works with data has heard of Edward Tufte, the guru of data display. Tufte is Professor Emeritus of Political Science, Statistics, and Computer Science at Yale University. He has published four books on various aspects of data display, and he also travels the country giving one-day courses on data display.

What makes Tufte such a prominent figure in this field is that he sets high standards for how data should be presented to the audience. In his books and courses he draws from charts and graphs from history to illustrate principles of good data display. A notable example is Charles Joseph Minard’s 1869 data-map of the losses suffered by the French army in the course of its invasion of Russia in 1812 (Tufte, 2006). As Tufte puts it, “vivid historical content and brilliant design combine to make this one of the best statistical graphics ever” (2006, 122).

Beautiful Evidence is Tufte’s fourth book. Its focus is on how evidence should be displayed. In his words, “how seeing turns into showing, how empirical observations turn into explanations and evidence. The book identifies excellent and effective methods for showing evidence, suggests new designs, and provides analytical tools for assessing the credibility of evidence presentations” (9).

The book has chapters on many of data displays, but two are of particular interest to us who work with educational data. One is his fifth chapter, The Fundamental Principles of Analytical Design. In this chapter, Tufte uses Minard’s data-map of the French army to illustrate six fundamental principles of analytic design:

- **Principle 1: Comparisons**: Show comparisons, contrasts, differences.
- **Principle 2: Causality, Mechanism, Structure, Explanation**: Show causality, mechanism, explanation, systematic structure.
- **Principle 3: Multivariate Analysis**: Show multivariate data that shows more than 1 or 2 variables.
- **Principle 4: Integration of Evidence**: Completely integrate words, numbers, images, diagrams.
- **Principle 5: Documentation**: Thoroughly describe the evidence. Provide a detailed title, indicate the authors and sponsors, document the data sources, show complete measurement scales, point out relevant issues.
- **Principle 6: Content Counts Most of All**: Analytical presentations ultimately stand or fall depending on the quality, relevance, and integrity of their content.

As one who works with data quite often, I was very interested in these principles, especially the extent to which they overlap with my own principles for data display. In my experience in our field, audiences tend to have varying levels of comfort with data. As a result, I try to present data, at least initially, as simply as possible: one, then two, but no more than two, variables at a time, as necessary to answer the primary analytic question. Viewers inevitably raise questions pointing to additional variables to consider. In response, I try to be prepared with multivariate charts that include these additional variables. But complicated charts that try to include too much variation lose people. Tufte holds that people really do want a more complex multivariate display that tells more stories and captures more of the complex reality being investigated. Duly noted, Dr. Tufte, but my experience urges caution here.

I find myself agreeing with most of Tufte’s other principles. Still, I wonder: Are these really universal principles, or just Tufte’s preferences for good data display? (For we all have preferences here, don’t we? For example: Many people like pie charts; I can’t stand them.) So as I read, I wondered if there is a body of controlled scientific research on which displays of data (tables, line graphs, bar graphs, etc.) are more cognitively effective than others to people in various settings. If there is, I didn’t see Tufte cite it. Rather, he contends that these principles are universal on the ground that they can be found in data displays throughout history and cultures. Well, I’m not sure I buy the universality of Tufte’s principles, but they do make a lot of sense. I think he is right to insist that data displays do justice to the content or research question. In our age of data dashboards, it is easy to get caught up in fancy data displays and lose sight of the theory, hypothesis, or fundamental research question that the data are supposed to answer. I know I can do more to integrate words, numbers, and images, as well as to properly document measurement scales, data sources, and authorship. That seems like good practice.

Another important chapter is the seventh, The Cognitive Style of PowerPoint (which is also available for download (for a fee) from his Web site at www.edwardtufte.com). Tufte is very critical of PowerPoint on a number of grounds. As he summarizes it well on page 158:
PowerPoint’s convenience for some presenters is costly to the content and the audience. These costs arise from the cognitive style characteristic of the standard default PP presentation: foreshortening of evidence and thought, low spatial resolution, an intensely hierarchical single-path structure as the model for organizing every type of content, breaking up narratives and data into slides and minimal fragments, rapid temporal sequencing of thin information rather than focused spatial analysis, conspicuous chartjunk and PP fluffy branding of slides with logotypes, a preoccupation with format not content, incompetent designs for data graphics and tables, and a smirky commercialism that turns information into a sales pitch and presenters into marketers. This cognitive style harms the quality of thought for the producers and the consumers of presentations.

These points too are well taken. As someone who has both consumed and produced innumerable PowerPoint presentations of varying quality, I can definitely attest to these limitations. Its bullet-point structure is not very well suited to lengthy explanations or well-developed arguments. I find its charting tool clunky for all but the simplest of charts. To go for more sophisticated charts, or to shrink fonts too small in order to include more detail is to sacrifice visibility to viewers farther from the screen. That can be frustrating.

Still, I think we can overcome some of these issues to use PowerPoint reasonably well. For my own part, I skip nearly all of the PowerPoint auto-formatting and just work with blank slides. I use Excel to generate charts and tables and then paste them into slides. I use my own text boxes to add any necessary text. No distracting clip art, photos, or animations. I also try to be very clear and forthright about the purpose and overall organization of the presentation, and I strive to make all slides work together to build an overall argument or narrative.

Beautiful Evidence is a readable and thought-provoking book which I recommend to educators who regularly use quantitative data as evidence. It is definitely rich with visuals, and I found Tufte’s writing economical and to-the-point. Tufte’s work has set standards of quality that I try to live up to. I don’t always succeed. Alas, not everything I do is Minard’s data-map of the march of the French army!

References


--Jack Monpas-Huber is Director of Assessment and Student Information for Shoreline Public Schools. He is a frequent contributor to WERA publications. Contact him at jack.monpas.huber@shorelineschools.org.
If you wanted to create something directly opposed to how the brain functions best in business, you might invent a cubicle. In education? How about a classroom? John Medina thinks we need to start over and he has 12 brain rules we can follow.

Medina is a developmental molecular biologist, research consultant, affiliate Professor of Bioengineering, director of the Brain Center for Applied Learning Research at Seattle Pacific University and self-proclaimed “grumpy scientist.” The research stated in his book has to pass the MGF (Medina Grump Factor) meaning that it is peer-reviewed and successfully replicated. However, rather than fill his book with an extensive list of research, Medina offers a website (www.brainrules.net) where those who wish may access it.

True to Brain Rule #10 (Vision trumps all other senses), Medina enhances his website with a series of short, focused videos which introduce each brain rule. I found these clips extremely useful in sharing my epiphanies with my husband who is no old-fashioned-soak-in-the-tub-and-read-until-pruny-book-lover (see Rule #3: Every brain is wired differently). I showed him a 3-minute video from the site and we had great context to discuss Medina’s findings.

I read this book as part of a woman’s book club that chooses books that will help us move in new directions in our lives. This book delivered. The brain rules that had the most impact on changing my own thinking and behavior were Rule #1 on exercise, Rule 4 on boring things and Rule #7 on sleep.

On the subject of exercise Medina writes: “All of the evidence points in one direction: Physical activity is cognitive candy. We can make a species-wide athletic comeback. All we have to do is move.” Medina’s mantra “all we have to do is move,” has changed the way I make time for my need to be active. It reminds me of Sir Ken Robinson’s quip about how our bodies were not designed to carry our heads from meeting to meeting. So now, when I have a problem at work, I don’t sit at my desk. I get up and walk around the building; I walk downstairs to the mailroom and back; I invite a colleague to walk. When things don’t make sense, movement always does.

In my work as an instructional coach, I have the honor of working with students, teachers and administrators at thirteen different elementary schools. Whether principals are talking about something that was said at a staff meeting or teachers talking about kids retention of what they taught yesterday, the question is the same “If we taught it/said it. Why don’t they remember?” An important part of the answer is rule #4: We don’t pay attention to boring things.
Medina likes to ask his students when, in a class of even medium interest, they start glancing at the clock, wondering when the class will be over. Inevitably, one nervous student comes up with the response (shared by most) that this happens after about ten minutes. Indeed, research confirms that after ten minutes, we need to talk, move, and shift in order to keep our attention.

In regards to rule 7 (Sleep well, think well), after exercising or doing one thing for awhile, I'm ready for a siesta. I've always thought I needed to live in a country that believes in shutting down in the afternoon and resting. I was delighted to read that "people vary in how much sleep the need and when they prefer to get it, but the biological drive for an afternoon nap is universal."

Medina shares different chronotypes that exist in our society. Larks, for instance, often rise without an alarm clock by 6 a.m., claim to be most alert around noon and most productive a few hours before lunch. Owls on the other hand rise only with the help of an alarm, would prefer to doze until noon, and work productively into the late evening. While naturally an owl, I've learned to act more larkish, with a sleep schedule that aligns better with the typical work day. Still, I ponder, as does Medina, how we might organize our days differently to acknowledge varied chronotypes and increase productivity.

John Medina’s grumpy science is equally balanced with his delightful sense of humor and personal anecdotes. At the close of the book, I found myself hoping that we might all take the time to rewire and follow brain rules more closely.


--Heather Rader is an instructional specialist for North Thurston Public Schools.
Contact her at hrader@nthurston.k12.wa.us.
My Favorite Kosher Vegetarian Punjabi Restaurant
Reviewed by Andrea Meld, Ph.D.

You might just drive by and never notice it. Pabla Indian Cuisine Restaurant, an oasis of lace curtains and fragrant spices, is located in the Renton Fred Meyer Complex, at 364 Renton Center Way SW, Suite #C60. A family-owned restaurant that first opened in 1998, Pabla serves a diverse menu of spicy and savory vegetarian Punjabi food. The family restaurant business goes back to 1947 in India.

Pabla Indian Cuisine does not use any meat or eggs, and does not serve alcoholic beverages. On request, most of the dishes can be made “vegan” that is, without using any dairy products. Pabla Indian Cuisine also has kosher certification from the Va’ad HaRabanim of Seattle, in response to an outpouring of requests from the Jewish community in nearby Seward Park. Vegetarian food, either with or without dairy products works very well for those adhering to kosher rules about food, which prohibit having or mixing dairy and meat products at the same meal. But the most important reason for dining at Pabla Cuisine is that the food is absolutely delicious and made with fresh and healthy ingredients.

The Punjab, a lush region in Northern India, is one of the most productive centers of agriculture in the world, and supplies much of India with wheat, rice, other grains, sugar and fruit. Spices that are liberally used in Punjabi cuisine include ginger, garlic, cardamom, chili peppers, black pepper, cinnamon, cloves, cumin, coriander, bay leaf, and garam masala, a hot spice mixture. A formal dinner might consist of one or two appetizers, a salad, some form of bread, a rice dish, a lentil dish, a curried or sautéed vegetable dish, and some form of paneer, a home-made cheese, especially if the meal is vegetarian. Condiments might include raita, a yogurt dish, and some kind of pickle or chutney. A dessert: Try some gulab jamon, a confection made with rose essence, saffron, cardamom, and almonds, or perhaps you would prefer some homemade mango ice cream. If you decide to pass on dessert, be sure to have a cup or two of hot spiced chai tea. Savor and enjoy.

For about $10, you can enjoy a buffet lunch, served from 11:00 am to 3:00 pm, and sample from over 20 items. Pabla is an excellent choice in value and time if you are attending either or both of the Spring WERA events at Puget Sound ESD in Renton, or if you prefer to dine in a more leisurely way, you can order from the dinner menu. Here are some of my suggestions.

Appetizer: Vegetable samosas, chile pakoras, or samosa chat (two samosas with mint sauce, tamarind sauce, plain yogurt with cilantro). Samosa chat is one of the best appetizers ever.

Special tandoori breads: Nan, a leavened bread baked in a tandoor (clay oven), paratha, a buttered whole wheat bread, roti, a whole wheat bread somewhat like a tortilla, or puri, a fried bread that is puffs up to a hollow ball. All are tasty.

Soup: The Pabla tomato soup is excellent and somewhat lighter than the Pabla special soup, made from lentils, cream, and garlic, also good. The Daal (lentil) soup is also quite good. If you prefer salad to soup, I would highly recommend the Punjabi Indian salad, which consists of tomato, lettuce, cucumber, onion, green chilies and a tangy dressing.

Entree: I haven’t tried all of the entries yet, however, all that I have tasted are delectable. Be sure to have dish containing paneer, the homemade cheese, prepared with curry sauce and spices, a dish made with eggplant, and some saag channa, a dish made with spinach and chick peas, and some type of potato or cauliflower dish. You might also prefer a korma dish, which are somewhat creamier and sweeter. If you’re not sure, the wait staff will be happy to make suggestions.

Dessert: Try some gulab jamon, a confection made with rose essence, saffron, cardamom, and almonds, or perhaps you would prefer some homemade mango ice cream. If you decide to pass on dessert, be sure to have a cup or two of hot spiced chaï tea. Savor and enjoy. Namaste.

Pabla Indian Cuisine
364 Renton Center Way SW #C60
Renton WA 98057
Phone: (425) 228-4625
www.pablacuisine.com/main/site/index.html

References


--Andrea edits The Standard Deviation and the Journal’s ethics column. She is an OSPI data analyst. Contact her at Andrea.Meld@k12.wa.us.
Reviewer Acknowledgement
Volume 3 The WERA Educational Journal

David Anderson, Office of Superintendent of Public Instruction
Jan Bakken, Everett Public Schools
Pat Cummings, Tacoma Public Schools
Phil Dommes, North Thurston School District
Janet Fawcett, Renton School District
Michael Flicek, Natrona County (WY) Schools
Yoonsun Lee, Seoul (Korea) Women’s University
Duncan MacQuarrie, Council of Chief State School Officers
Brian Rick, Bellingham School District
Marcy Stein, University of Washington Tacoma
David Thielk, Office of Superintendent of Public Instruction

Many thanks to the critical friends of WERA who have helped make The WERA Educational Journal a (partially) peer reviewed journal. See the guidelines for peer reviewers in the following pages. --Editor
Manuscript Review Form
for use when evaluating research manuscripts

MS Title ______________________________________________________________ _

MS# ___________ Reviewer ID# ___________ Review Due Date ____________

I. CRITERION RATINGS
Please circle the appropriate rating for each criterion.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The overall soundness of the approach (e.g., research methodology,</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>logic of analyses, etc.) .....................................................................</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. The practical relevance, use or other value of this article to those</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>working in Washington schools and colleges ......................................</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. The overall soundness of the inferences, conclusions and implications</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>for practice ......................................................................................</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. The quality and timeliness of the references cited ..........................</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>5. The organization, writing quality, and readability of the manuscript.</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>6. Adherence to the journal’s submission guidelines ............................</td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>

II. NARRATIVE ASSESSMENT
On a separate sheet of paper bearing (a) your Reviewer ID# and (b) the MS# (see above), please type a narrative assessment of the manuscript, emphasizing its strengths and weaknesses. If you are recommending outright rejection, please state the major reasons. If you are recommending acceptance or revisions, please provide specific information that will help the authors improve this piece. In most cases your assessment will be shared with the authors; please keep that in mind when writing.

III. SUMMARY RECOMMENDATION
Please check one
_____ Accept (require minor or no revisions)
_____ Conditionally Accept (require substantial revisions)
_____ Reject/Encourage (resubmission after revisions)
_____ Reject (do not encourage resubmission)

IV. DATE OF REVIEW: ________________

V. DECLINE TO REVIEW BECAUSE: ___________________________________________

Please return review via U.S. mail, fax, or e-mail (preferred) to:
The WERA Educational Journal
Peter Hendrickson, Editor
Union Hill Grassfed Beef, 7550 205th Ave. NE, Redmond, WA 98053
Fax: (425) 868-8318
E-mail: p.hendrickson43@gmail.org
Phone: (206) 658-5930
(Rev./ph/11mar2011)
Research, Program Evaluation, Assessment and Policy Manuscripts Review Form

Instructions for Reviewers:

This form has been designed to help you judge the suitability of the attached manuscript for publication in The WERA Educational Journal. The editor will use the information you provide on this form (along with similar information provided by other reviewers) to decide whether this manuscript should be: (a) accepted with minor or no revisions, (b) conditionally accepted with substantial revisions, (c) rejected with resubmission encouraged, or (d) rejected outright.

We are not a first tier journal. Many authors have been invited to submit manuscripts. For a few, this may be their first venture into academic writing for publication. In some cases, an editor has already reviewed the manuscript to help the author make it more suitable for submission. A goal of the Journal is to encourage academic writing among members. Unlike many of the state or regional AERA affiliates, WERA is largely composed of public school teachers, administrators and other staff. Many present but few write as scholarly productivity is not an evaluation criterion for most.

Please complete all sections of this form and attach a separate sheet with your typed narrative assessment. In completing this review form, please keep the following points in mind:

(a) If you know that you will be unable to meet the review deadline, please return the manuscript to the editor immediately, indicating your inability to review.

(b) As we are a smaller public school/university academic community, we have not removed author(s)’ identification from the manuscript. If you think that knowledge of the author’s identity will impair your ability to assess this manuscript in a fair and unbiased manner, please disqualify yourself as a reviewer of this manuscript and indicate your inability to review.

(c) In completing your review, please do not concern yourself with minor technical imperfections in the manuscript. Violations of APA style and minor writing problems will be handled by the journal editors after the more important, substantive review is completed. Do not edit the manuscript itself and do not return it to the editor; except in extraordinary circumstances, manuscripts will not be returned to authors. Instead, please include all comments on this form and in the accompanying narrative assessment.

(d) The first section of the review asks you to rate the manuscript with respect to six criteria. Although the individual ratings will contribute to the final editorial decision, these ratings will not be tallied to produce any kind of overall score.

(e) Thank you for your prompt and conscientious work on behalf of the journal and, more broadly, on behalf of public education in the Northwest.
The WERA Educational Journal

Editor
Peter Hendrickson, Ph.D.
Research and Evaluation Consultant
206-658-5930
p.hendrickson43@gmail.com

Book Review Editor
Phil Dommes, Ph.D.
North Thurston School District

Food Editor
Heather Bandeen, Ph.D.
OSPI

Ethics Editor
Andrea Meld, Ph.D.
OSPI

Humor and Puzzles Editor
Jeff Katims
OSPI

Photo Editor
Don Schmitz
Mukilteo School District

Technical Editor
Patrick Cummings,
Tacoma School District

Editorial Assistants
Jeanne Willard &
Loretta Comfort
Everett Public Schools

Layout Designer
Michelle Sekulich
The Boeing Company

Executive Secretary
Sharon Rockwood
WERA

The WERA Educational Journal is published fall and spring as an occasional peer reviewed online journal. Submissions are welcomed from WERA members and others. Kindly submit articles for consideration using APA format. Copy deadline is September 23rd, 2011.