THE STANDARD DEVIATION

February 2009 Newsletter and Educational Journal

Washington Educational Research Association
University Place, WA

http://www.wera-web.org

WERA 2009 Spring Conference
Returns to Airport Hilton

The Washington Educational Research Association

Spring 2009 Conference

Between the Bookends: Connecting standards-based instruction to the research on assessment.

March 26-27, 2009
(Pre-conference March 25)

Seattle Airport Hilton Hotel Conference Center
17620 Pacific Highway South
Seattle, WA 98188

WERA's Spring conference will feature two nationally known keynote speakers and two days of breakout presentations from educators across the state. There will be presentations of value for teachers, principals, curriculum specialists, central office staff, and post-secondary educators.

Thursday's keynote presenter is Dr. William Schmidt. William Schmidt is a University Distinguished Professor at Michigan State University and the National Research Coordinator and Executive Director of the National Center which oversees U.S. participation in the Third International Mathematics and Science Study (TIMSS). A past Chairman of the Department of Educational Psychology and former Acting Dean for Planning and Evaluation in the College of Education at Michigan State University, he was also head of the Office of Policy Studies and Program Assessment for the National Science Foundation.

Friday's keynote presenter is Larry Ainsworth, Executive Director of Professional Development at Doug Reeves' Leadership and Learning Center in Englewood, Colorado. He travels widely throughout the United States to assist school systems in implementing best practices related to standards, assessment, and accountability across all grades and content areas. He is the author or co-author of eight published books: *Power Standards, "Unwrapping" the Standards, Common Formative Assessments, Student Generated Rubrics,* and *Five Easy Steps to a Balanced Math Program,* including three 2006 editions, one each for the Primary, Upper Elementary, and Secondary grade spans. Larry's primary motivation is to assist educators and leaders in helping all students succeed by "taking the mystery" out of the instruction, learning, and assessment process.

Wednesday's pre-conference sessions will feature in-depth information on topics related to the conference theme. Of particular interest may be the pre-conference session on standards-based grading. Several districts currently implementing standards-based reporting will be sharing their work. On Thursday and Friday, the conference will offer more than 30 breakout sessions.

The conference includes the Pete Dodson Symposium, which promises to be a stimulating discussion related to the question of **"State, District, and Local assessment… how much is too much?"** With our recent election focusing on the burden the WASL places on educators and students, this promises to be a lively discussion. Gordon Ensign, former Director of the Commission on Student Learning and a past WERA president, will serve as moderator.

Please register now for an exciting conference. As always, free clock hours will be available. www.wera-web.org

(Continued on next page…)

INSIDE THIS ISSUE

Wednesday, March 25, Pre-Conference Workshops

This year's pre-conference program offers two all-day and 8 half-day training sessions. Attend one of the all-day sessions or choose one half-day workshop in the morning and one in the afternoon. *Pre-registration for specific sessions is required*. Continental breakfast and check-in begin at 7:30 a.m. *Morning workshops are 8:30 to noon. Afternoon workshops are 1:00 to 4:30. Lunch is scheduled from 12:00 to 1:00.*

Pre-Conference Workshop Descriptions

1. (All Day) Professional Learning Communities and Response to Intervention: Implications for District-Wide Implementation (Mike Jacobsen and Janel Keating, White River School District) The PLC journey will be described and show how PLC has provided the foundation for implementing an RTI model. Participants will learn how the three big ideas of a PLC – a focus on learning, a collaborative culture and a focus on results, make PLC and RTI natural partners. Information will be provided on implementation of collaborative teams, common assessments, power standards, universal screening, progress monitoring, a three tiered intervention model and data collection and display procedures.

2. (All Day) Interpreting Test Score Trends and Gaps (Andrew Dean Ho, University of Iowa) The reporting of proficiency rates (proportion of students above a preset cut point) is ubiquitous in state and federal accountability systems. Disaggregation of data is often required to focus on performance gaps in groups of students. Proficiency rates offer only a limited viewpoint. Organized by Northwest Regional Lab, this session will assist attendees in understanding limitations and proper use of proficiency rates, and will offer additional methods for using the data to support sound inferences and decisions.

3. (A.M.) Connecting Standards-Based Instruction and District Assessment Data to Improve Student Achievement in Literacy (Cindy Foster, Everett Public Schools) Everett Public Schools committed to a professional development model with literacy facilitators in every secondary school. This model uses assessment data to inform teachers about the effectiveness of their instruction to create curriculum that improves student learning. Facilitators customize real-time professional development by addressing the strengths/weakness found in the district student assessment data. Feel free to bring your own data to develop your own action plan and discover how this model could benefit your school/district.

4. (A.M.) Moving from Analyze to Adjust: Tools that Help Analyze Student Work to Improve Instruction (Tamara Smith, Dan King, and Robin Henrikson, Olympia ESD) This session will provide participants with tools, strategies, and resources to engage professional learning communities in analyzing student work and refining/improving instruction based on the results. Tools, student work samples, and strategies offered will be in mathematics, but are readily adaptable to all content areas.

5. (A.M.) College Readiness Mathematics Standards and the New College Readiness Mathematics Test (Russ Killingsworth, Seattle Pacific, and Kristen Maxwell, ESD 105) This interactive session is designed to engage participants in the College Readiness Mathematics Standards and their connection to the newly developed College Readiness Mathematics Test. Information regarding the development process of the test will be shared, current test implementation status will be unveiled, and opportunities for questions will be given.

6. (A.M.) The Missing Component in School Reform: Standards-Based Grading & Reporting (Tammy Campbell, Mary Weber, and Kathy Williams, Spokane Public Schools) Without implementing standards-based grading and reporting, standards-based instruction and assessment have less impact on student achievement. Spokane Public Schools will share its work in implementing standards-based grading at all of its elementary schools. This interactive session will showcase products we have created and allow time for participants to share standards-based grading & reporting materials they are using in their districts.

(Continued on next page...)

7. (P.M.) Secondary Standards-Based Grading: Turning Theory into Practice (Forrest Clark, North Thurston Public Schools) You have read books and attended workshops on standards-based grading. But how do you turn theory into actual secondary classroom practices? Learn how one district is addressing classroom and building-wide issues while turning theory into practice. This session will include examples of assessments, grade book entries, and progress reports.

8. (P.M.) The Flexibility of Peer Learning Labs (Jerry Johnsen and Jennifer Chase, Spokane Public Schools) When engaged in daily teaching we have few opportunities to listen intently or observe individual students or groups. Learn how Peer Learning Labs are tailored to the specific needs of the participants, and can be easily written to accommodate lesson study, exploration of a teaching dilemma, content teaming, and/or vertical teaming.

9. (P.M.) Growth Models for Classroom Assessment and NCLB (Joseph Stevens, University of Oregon) This session will discuss using longitudinal growth models. It will cover: a) analysis of change, b) common growth models with examples, c) research design for school/program evaluation, d) classroom applications of models, e) technical and statistical issues in growth modeling. Participants will have opportunities for discussion and application of the information covered.

10. (P.M.) Common Formative Assessments with Standards-Based Reporting (Nancy Coogan, Aaron Mukai and Laura Phillips, Mukilteo School District) This session will focus on the importance of short assessment cycles to increase student achievement based on Larry Ainsworth's work. Content will include both *mathematics and literacy.* Complexities revolving around how to create a culture where staff sees the benefit of short assessment cycles will be discussed.

Edmonds Woodway Jazz Choir directed by Charlotte Reese, performs at the Winter State Assessment Conference

President's Column

The WERA board continues to seek ways to encourage academic discourse about educational issues during and through our conferences and events. The December Assessment Conference, ***Assessment Dynamics for Dynamic Learning***, was well-attended with 377 total participants. The conversations about state, district, and classroom assessments were rich and collegial learning was applicable to our educational settings.

WERA memberships are similar to the last two years. As of today we have 436 current members. There is still time to rejoin for the 2008-09 year (http://www.wera-web.org), if you haven't done it yet. Elections for president-elect and two at-large board members are in February, and only current members received ballots. You will find the biographical information on all candidates elsewhere within this issue of *The Standard Deviation*.

The WERA board has made a conscious effort to develop a frugal budget for 2008-2009. Much like districts and other public entities, we looked to maintain our modest $25 membership fee which is returned to members through research grant opportunities, internet services, and publications. We also tried to balance rising hotel fees by eliminating the evening hospitality session, dropping give-away items, and handling conference evaluations online. We will be sharing other cost-saving measures at the annual membership meeting which will be held during the WERA spring conference in March.

There are two major WERA events coming soon. Past President Lorna Spear is heading up the conference planning committee for the WERA Spring conference, March 25-27 at the Seattle Airport Hilton. A multitude of timely sessions will offered, as well as two keynote presentations on the topic of connecting standards-based instruction to the research on assessment. A single topic workshop on formative assessment with Dylan Wiliam is scheduled for June 25, 2009 at the Puget Sound ESD in Renton. You can sign up online for either offering at the WERA website.

In this time of uncertainty in the realm of educational funding and shifting program mandates, such as reauthorization of No Child Left Behind, you can be certain that WERA will remain a prominent force in supporting educators and students around the state of Washington. We invite you to join us in ongoing discussions about the changing landscape of education and in providing a voice to shape educational policy and practice for many years to come.

-Nancy Arnold, Ed.D. is Director of Special Programs for Puyallup School District and WERA President. She was a special education assessment specialist with OSPI.

The mission of the Washington Education Association is to improve the professional practice of educators engaged in instruction, assessment, evaluation, and research.

WERA Services

- *WERA provides professional development through conferences, publications, and seminars.*
- *WERA provides forums to explore thoughtful approaches and a variety of views and issues in education.*
- *WERA provides consultation and advice to influence educational policy regarding instruction, assessment, evaluation, and research.*

Executive Board Slate of Candidates for 2009-10

The Executive Board has voted to present the following slate of candidates to the membership. The WERA Constitution calls for an election in February of each year. WERA members were invited at the WERA/OSPI Assessment Conference in December to nominate other candidates. No additional names of candidates for any office were offered.

On the accompanying pages you will find a brief biography of each candidate, along with their statement of vision for WERA. Printed ballots will be mailed to members in February. Please make sure your WERA dues are current for 2008-09. Ballots will be sent to members only.

All positions require three year commitments from candidates. For the two at-large positions, there are three candidates for each position. You will be asked to rank the candidates for each position. If no candidate receives a majority on the first ballot, the candidate receiving the least support will be eliminated, and a recount will be done between the two remaining candidates using all ballots for only the two remaining candidates.

Winners for these Board positions will be announced at the WERA Spring Conference in March.

President Elect

Gene Sementi
Assistant Superintendent, West Valley School District, Spokane

I have worn a lot of different hats in my twenty two year career in education including teaching math and science in middle and high school, coaching at the middle and high school levels, serving as a middle school assistant principal, an elementary principal, a middle school principal, and as a high school principal. Presently I am serving as an assistant superintendent for the West Valley School District in Spokane. In addition I have worked extensively with the Association of Washington School Principals as an instructional leadership and data driven decision-making consultant, in an effort to assist struggling schools. I am still a teacher at heart and am teaching, or have taught, Ed. Law at EWU, Instructional Leadership at Whitworth College, and Education Leadership at WSU to name a few. I was awarded the Washington State Sharon Christa McAuliffe Award for Excellence for my school improvement work while I was the principal of Orchard Center Elementary School. I was also recognized as the Washington State Middle Level Principal of the Year for my work at Centennial Middle School, and I have recently been awarded the Washington State Association for Supervision and Curriculum Development statewide individual award for Pursuing Continual Lifelong Learning.

My first involvement in education research was in my mathematics classroom where I sought to determine the early predictors for student success in upper level high school mathematics. Additionally my major research experience was through my doctoral dissertation work at the University of Idaho, where I sought to pinpoint the personal, professional, and affective attributes that, when identified through the hiring process, would identify the candidates most likely to become highly effective teachers.

I am a long time WERA consumer where my first involvement was attending a conference with my school's math improvement team. The team and I quickly came to realize the value in the nuts, bolts, and how-to-do-its provided at WERA conferences. In the years since I have attended, with a team from my district, at least one WERA conference a year for the past several years. Recently I served as an at large board member for WERA and have participated, in some small way, in the behind the scenes work that has helped to make WERA one of the most respected Educational Research Associations in the nation.

(Continued on next page...)

President Elect *(continued)*

Bob Silverman
Executive Director of Assessment and Accountability, Puyallup School District

I have been an educator in Washington since 1995. Throughout this time period I have been an active participant in WERA, serving as both a Board Member and as President. Additionally, I have chaired, co-chaired, or served as a conference committee member for many of WERA's annual conferences and training sessions. Currently, along with three statewide colleagues, I co-chair WERA's Assessment Directors' Network. WERA has a long and distinguished history of advocating for educational research, assessment, and school improvement in our state. I am proud to have been a contributor to that work and am eager to extend that participation.

As Executive Director for Assessment and Accountability for the Puyallup School District, I am responsible for leading our district's efforts in assessment and research; and for helping our schools respond to state and federal accountability requirements. Our goals in our district mirror well those of WERA. The statewide leadership that WERA has provided has benefited our district greatly. I would look forward to being able to help participate in and shape WERA's future.

Member-at-large Position 1

Jack Monpas-Huber
Director of Assessment & Student Information, Shoreline Public Schools

As Director of Assessment & Student Information for Shoreline Public Schools, I am committed to helping that system realize its vision of academic excellence for all students through improved decision-making based on valid and reliable information. Prior to joining Shoreline I was Director of Assessment & Program Evaluation for Spokane Public Schools, and before that I served Northshore School District for six years as Assessment Data Analyst. I am a member of the American Educational Research Association (AERA), the National Council on Measurement in Education (NCME), and the National Association of Test Directors (NATD). I hold a B.A. in Sociology from Arizona State, a M.S. in Sociology from Virginia Tech, and a Ph.D. in Educational Psychology (Measurement) from the University of Washington.

WERA has been, and continues to be, a valuable source of professional development for me. At WERA conferences I have always gained valuable information that I could integrate into my own practice. They have also connected me with colleagues in other districts who share my interest in applying research and data to the real challenges of practice. My vision for WERA is to promote strong, powerful connections between theory, research, policy, and practice in two ways. One is by actively recruiting a diverse membership of people working in different settings who share an interest in educational research. The other is by actively encouraging and supporting educational research of various kinds. To bring these two together could create a rich source of professional learning that, when applied in schools and classrooms, could improve the quality of education for all students in Washington State.

Brian Rick
Assessment and Evaluation Specialist, Bellingham School District
During my 16 years as an educator I have had many opportunities to research, create, report, evaluate, use (and misuse?) data. I have been a technical and community college instructor, high school math teacher, TOSA, and assessment specialist. My formal preparation came from WWU, earning a BS in Math and MEd in Secondary Education. My experiences have allowed me to assist with various groups, including the state's Math Assessment Leadership Team and WASL committees for math content, test spec. reviews, data reviews and range-finding. I also had the privilege of supporting teachers in the Whatcom Skagit Mathematics Partnership as internal evaluator. Currently I serve on the State Technical Advisory Committee, CAA Options Technical Advisory Committee, and the Alternate Assessment Advisory

(Continued on next page…)

Member-at-large Position 1 *(continued)*

Council. In my work I have certainly been assisted by my membership in WERA, especially as a participant and presenter at WERA and OSPI and assessment conferences.

With that background, I am ready to ask for a new role in WERA as a Member-at-Large on the Executive Board. WERA continues to highlight essential educational issues we need to address, promote research we can use, and provide opportunities for us to learn. As the needs of the WERA membership grow with its increasing diversity, I would like to lend my efforts toward expanding professional development opportunities and encouraging research projects that support the practitioners in our schools.

Kathryn Sprigg
Assistant Director, Office of Accountability, Highline School District

Education is in my blood; my mom and grandma were teachers, as are two of my children. As the Assistant Director for the Office of Accountability in the Highline Public School District, I work to facilitate schools' use of data to inform their instructional practice and improve student learning. One of my goals is to find ways to streamline our processes and help schools find more time to devote to teaching.

My participation in WERA has provided me with information and research that helps me think about how I can use our own research and data to accomplish this goal.

WERA is an important resource for me in my professional life and it's an important partner for districts and the state as we continue our commitment to high standards for everyone involved in Washington's education system. I've attended many conferences, presented at several of them, encouraged my colleagues to join WERA and attend the conferences, and volunteered on conference planning committees. I'd honor the opportunity to increase my involvement and contribution by serving on the WERA Board.

Before joining Highline, I worked at OSPI, where I was the State Coordinator for the National Assessment of Educational Progress (NAEP). I earned my doctorate from Seattle University in Educational Leadership and Policy Studies, and a post-doctoral certificate in Large Scale Assessment from the University of Maryland, College Park. I've presented my research at AERA, CCSSO and the National Assessment Governing Board, among other venues. The experience I've gained at the national level will help me provide perspective to the WERA Board as they make decisions about our direction and goals for the future.

Member-at-large Position 2

Ryan Grant
First Grade Teacher/Program Coordinator, Medical Lake School District

I teach first grade at Michael Anderson Elementary School, a Pre-6 building located at Fairchild Air Force Base outside of Spokane. All of my 8 years of teaching have been in first grade; my passion is for teaching reading, and watching the growth they make in first grade is the best reward any teacher could ask for. At Anderson I serve on the pre-referral team (Success) for kids who are struggling, as well as coordinating the MAP assessment for grades K through 6 and serving as the liaison for the NAEP test that we'll be a part of this year. In 2007 I was awarded a WERA research grant to study how elementary schools can best implement the principles behind Response to Intervention, and I'll be presenting the results of that research at this year's WERA Spring Conference.

(Continued on next page…)

Member-at-large Position 2 *(continued)*

As a classroom teacher I'm most interested in exploring ways to make the results of research practical for the classroom teacher to use. This is a theme that I'd love to explore more through WERA; how to close the research-to-practice gap, why that matters, and why educational research (particularly teacher-directed action research) is not only nothing to be afraid of, but also has the promise to positively and completely remake our schools. WERA has already been a leader on this front, and I'd love to help in the conversation.

Mike Jacobsen
Assessment and Curriculum Director, White River School District

I have been an educator in Washington for 28 years. Prior to my current role as Assessment and Curriculum Director I served as a special education administrator and as a school psychologist. In the role of a school psychologist, I was a Past President of the Washington State Association of School Psychologists and received the School Psychologist of the Year Award. In addition I was also recognized at the national level with a Legislative and Advocacy award from the National Association of School Psychologists of which I have also served as Western Regional Director and Ethics Chair. I have had the opportunity to serve as adjunct faculty at the University of Washington, Seattle and Tacoma campuses, Seattle University and City University. I currently serve on OSPI's, RTI Leadership and LDA Committees. I have also had the opportunity to be published in *School Psychology Review,* a peer-reviewed journal.

During the course of my career in education I have been a member of a number of different organizations. I have found WERA to be one of the most valuable of any with which I have been involved. My passion centers around using assessment data to improve instruction and student performance. I believe WERA's mission of improving: "the professional practice of educators engaged in assessing student performance, evaluating programs, conducting and applying educational research and using data to inform instructional practice" is critical to moving our educational system forward in Washington. I am honored to have presented at WERA conferences a number of times. I would encourage WERA to continue to offer high quality conferences, increase its support of local research and increase the association's ability to disseminate research findings.

Jim Leffler
Program Director, Services to the Field – Northwest Regional Educational Lab (NWREL)

Having served the last three years on the WERA Board, I am excited about continuing to work in that capacity. I am currently on staff at NWREL, working with schools, districts and state offices in Washington, Oregon, Idaho, Montana and Alaska. My background and experience as a test director in Washington has served me well in this role at NWREL. My experience as a school principal, elementary and middle school teacher (although we called it Junior High in those days), grants manager with an ESD, and curriculum specialist have also helped in my work at NWREL. Having retired from Washington with 30 years, I moved to NWREL seven years ago. The work with schools and the work in more formal research across the region have given me a much broader perspective, as well as a much greater appreciation for all of the work and accomplishments of Washington state. While my children are grown and out of the school system, my wife, a primary teacher, helps keep me "grounded" in the reality of the day to day work of schools.

I feel WERA is entering a new phase as an organization, just as the State Superintendent's role appears to be entering a new phase. The work and leadership of WERA will continue to be critical in helping schools do their work. In a time of challenges, as WERA strives to stay current in meeting the needs of a more diverse membership, I feel it is important that the Board have continuity of membership – and I would like to help provide a part of that continuity.

Workshop on Formative Assessment

SAVE THE DATE

Professional Development Opportunity
Washington Educational Research Association

June 25, 2009

Workshop on Formative Assessment With Dylan Wiliam. Currently the deputy director of the Institute of Education, University of London, Dylan Wiliam has been a teacher in urban London schools, a leader in higher education, and a senior research director in the research and development division of the Educational Testing Service. He is also the co-author of a very influential and widely -cited review of research evidence on formative assessment, "Inside the Black Box".

At this workshop Dylan will share insights on how to make greater use of assessments to support learning. He contends that, if we are serious about improving student achievement, we must focus on teachers' minute-to-minute and day-by-day use of assessment to adjust instruction. According to Dylan, this deep changes requires a different form of teacher professional development: building-based teacher learning communities.

Registration Will Take Place On-Line Only

Please register at www.wera-web.org
Visa or MasterCard required
Registration opens on April 1, 2009

- Six clock hours are available. Clock hours will be free.
- Registration limited to first 100 people.
- An e-mail confirmation will be sent to people who register.

Thursday, June 25, 2009
Location: Puget Sound ESD
800 Oakesdale Ave. SW
Renton, WA 98055

8:30 a.m. - 3:30 p.m.
$150.00 Registration Fee
*Lunch is included in fee.

Future Calendar

WERA Items

- 2009 Spring Assessment Conference,
  March 25-27, 2009
  Hilton Seattle Airport Hotel

- WERA Test Directors August Meeting
  August 6, 2009
  Location TBD

- 2009 State Assessment Conference,
  December 9-11, 2009
  Hilton Seattle Airport Hotel

Other Calendar Items (Non-WERA)

- American Educational Research Association, National Council on Measurement in Education, National Association of Test Directors, Directors of Research and Evaluation Annual Meetings and Conferences, San Diego, CA,
  April 13-17, 2009

- American Evaluation Association Annual Meeting, Orlando, FL,
  November 11-14, 2009

Results of Assessment Directors' Survey
By Phil Dommes, Ph.D.

Who manages the assessment load in districts across Washington and what are their challenges? These were a couple of questions the WERA Assessment Directors' Network hoped to answer in an online survey conducted last year.

Access to the survey was offered to all districts in Washington through an assessment e-mail list kept by OSPI. Individuals from 166 districts responded. The largest group of responding districts (43%) served fewer than 1,000 students, although 15 districts serving over 15,000 students were also represented. Assessment budgets ranged from under $500 for the 79 smallest districts to more than $500,000 for the largest.

27% of the respondents had less than 2 years experience in the assessment position; 21% had more than 10 years experience. 18% considered themselves novices in testing and measurement; 32% saw themselves as advanced. 40% were WERA members and, interestingly, 23% weren't sure. 65% had heard of the WERA Assessments Directors' Network, but only 34% had attended a meeting.

As one might expect, individuals charged with the assessment role had many job titles and a wide range of duties. Larger districts tended to assign individuals fairly narrowly to an assessment role, whereas smaller districts assigned the assessment role as one of many shared by an individual. Every respondent had responsibilities for managing or coordinating the WASL; the vast majority prepared, managed or analyzed data (81%), monitored AYP status (80%), prepared annual district reports (85%) and made board presentations (85%). A significant number also coordinated student learning plans (66%), conducted research (54%) and evaluation (59%), developed local assessments (48%) or scored them (35%).

Many of the districts gave benchmark tests in reading (about 3/4's of the elementary schools), math (2/3), and writing (1/3), and very few in science. In general, fewer benchmark tests were given at the secondary level. The same patterns exist for classroom-based assessments and for diagnostic assessments: more reading, less secondary. DIBELS, DRA, and MAP (NWEA), were the tests cited most often in questions about both benchmark and diagnostic testing. Another 10-20% of the respondents said their districts will be adding various tests this year.

In addition to answering the basic survey questions, respondents were also asked what their most pressing needs were, and how WERA or the network might respond. Input from this question (and from other survey comments) has been synthesized and used as the basis for discussions at the network meetings. Major areas of concern were providing support for newer assessment coordinators, improving communication about assessment issues, and general professional development.

-Phil Dommes is the director of Assessment and Gifted Programs for the North Thurston Public Schools and a WERA Board member.

Websites of Interest to Measurement Folks: Program Evaluations

http://ccsr.uchicago.edu
Consortium on Chicago School Research (CCSR) was created in 1990 after the passage of the Chicago School Reform Act that decentralized governance of the city's public schools. Researchers at the University of Chicago joined with researchers from the school district and other organizations to form CCSR with the imperative to study this landmark restructuring and its long-term effects. Since then CCSR has undertaken research on many of Chicago's school reform efforts, some of which have been embraced by other cities as well. Thus CCSR studies have also informed broader national movements in public education.

http://www.ascd.org/publications/educational_leadership/dec08/vol66/num04/The_New_Stupid.aspx
Frederick Hess' Educational Leadership article titled, "The New Stupid", claims educators have made great strides in using data. But danger lies ahead for those who misunderstand what data can and can't do.

(Continued on page 16...)

New! Index of Technical Articles from *The Standard Deviation*

By Andrea Meld, Ph.D.

As promised, we've started to index *The Standard Deviation*, to make articles of interest from past issues of The Standard Deviation just a few clicks away.

First to be indexed are the 'how-to' articles for software applications such as Excel, SPSS, Access and others, for the technically-savvy and those wishing to become more savvy. These articles have been written by WERA members, district assessment coordinators, analysts, specialists and others interested in sharing their software savvy.

As an example, some of you are already or will be working with transcript data that includes a course name and a course number in two separate SPSS variables. What if you need to combine the course name Algebra (string) with course number 101 (numeric) into one variable, so you have Algebra 101 (one string variable)?

Here is a quick syntax*:

```
DATA LIST LIST /course(A8) numb(F6).
BEGIN DATA
Algebra 101
Science 402
END DATA.
LIST.
STRING comb(A12).
COMPUTE comb=CONCAT(RTRIM(course),

LTRIM(STRING(numb,F4))).
LIST.
```

Resulting variables will be Algebra101 and Science402.

*Raynald Levesque SPSSTools

http://www.spsstools.net/Syntax/Concatenate/CombineStringAndNumber.txt

The more complete index of *The Standard Deviation* articles will be posted by fall, 2009. In the meantime, we encourage you contribute your software application tips and tricks, as well as other articles for consideration to *The Standard Deviation*.

-Andrea Meld, Data Analyst, Assessment and Student Information, OSPI, currently serves as a Member at Large on the WERA Board.

Outgoing Superintendent of Public Instruction, Terry Bergeson, recipient of the WERA Lifetime Membership Award

Index of Technical Articles for *The Standard Deviation*

Topic	Title of Article	Author	Issue	Link
Data Base	Finding a Useful Assessment Data Base	Bruce Kelley, PhD Battle Ground SD	Spring 2007, page 11	http://www.wera-web.org/links/TheStandard%205-21-07.pdf
Evaluation	Conducting Evaluations of Programs Which Have No Evaluation Plan	Michael A. Power, PhD Mercer Island SD	Winter 2007, page 18	http://www.wera-web.org/links/SD_January2007.pdf
EXCEL	Excel: Putting the Tool to Work (Vertical look-up)	Peter Hendrikson, PhD Everett Public Schools	Fall 2006, page 6	http://www.wera-web.org/links/TheStandard%2010-12-1.pdf
EXCEL	Excel: Using Pivot Tables	Pat Cummings, Tacoma Public Schools	Winter 2007, page 8	http://www.wera-web.org/links/SD_January2007.pdf
EXCEL	Stupid Excel Tricks for Assessment Folks (Duplicate records)	Pat Cummings, Tacoma Public Schools	Spring 2007, page 18	http://www.wera-wb.org/links/TheStandard%205-21-07.pdf
EXCEL	Stupid Excel Tricks: Macros	Pat Cummings, Tacoma Public Schools	Fall 2007, page 12	http://www.wera-web.org/links/TheStandard%2010-16-07.pdf
EXCEL	Stupid Excel Trick (Statistical functions)	Pat Cummings, Tacoma Public Schools	Winter 2008, page 44	http://www.wera-web.org/links/TheStandard%20Jan%202008.pdf
EXCEL	Stupid Excel Tricks (Working with date functions)	Pat Cummings, Tacoma Public Schools	Spring 2008, page 35	http://www.wera-web.org/links/TheStandard_5_2008.pdf
MODELS	Q & A on Logic Modeling	Kari Green, Open Program Evaluators Network	Winter 2008, page 35	http://www.wera-web.org/links/TheStandard%20Jan%202008.pdf
SPSS	SPSS: Putting the Tool to Work (SPSS for assessment)	Linda Elman, PhD Central Kitsap SD	Fall 2006, page 7	http://www.wera-web.org/links/TheStandard%2010-12-1.pdf
SPSS	SPSS: Exploring My WASL Data, Initial Scrubbing	Linda Elman, PhD Central Kitsap SD	Winter 2007, page 16	http://www.wera-web.org/links/SD_January2007.pdf
SPSS	Using SPSS to Identify HS Students Highest WASL Scores and other Functions	Linda Elman, PhD Central Kitsap SD	Spring 2007, page 22	http://www.wera-web.org/links/TheStandard%205-21-07.pdf
SPSS	Using SPSS Graphs to Visualize Your Data	Andrea Meld, Ph.D . OSPI	Spring 2007	http://www.wera-web.org/links/TheStandard%205-21-07.pdf
SPSS	SPSS Tips and Tricks and Beyond (Writing syntax, avoiding syntax errors, data display, resources)	Andrea Meld, Ph.D OSPI	Winter 2008, page 37	http://www.wera-web.org/links/TheStandard%20Jan%202008.pdf
SPSS & Surveys	Analyzing Surveys Using SPSS (Recoding and transforming variables)	Linda Ellman, PhD Tukwila SD	Fall 2008, page 20	http://www.wera-web.org/links/TheStandard%20100608.pdf
Surveys	Applied Survey Research	Sylvia Dean, Ed.D Evergreen Public Shools	Fall 2007, page 1	http://www.wera-web.org/links/TheStandard%2010-16-07.pdf
WEB	Comparing Schools Online: A Washington Guide	Peter Hendrikson, PhD Everett Public Schools	Fall 2007, page 7	http://www.wera-web.org/links/TheStandard%2010-16-07.pdf
WEB	Everything is on the Web: a Brief Review (Using on-line resources for research)	Kate Corby, Michigan State Univ. and Laura Lillard, UW, Seattle	Spring 2008, page 18	http://www.wera-web.org/links/TheStandard_5_2008.pdf

WERA Test Directors Meeting Notes

Test directors from across the state met during the OSPI/WERA Assessment Winter Assessment Conference and gave overwhelming support to a letter supporting a change in the legislation governing the requirement of high school students to WASL test annually if they have not met standard after 10th grade. They learned that a similar initiative is afoot with OSPI and are hopeful that students who did not test a second (or third) time will not be held to a higher standard—must pass—that those who tested. Nancy Katims of Edmonds drafted the petition for consideration by the 50 or so gathered. Test directors also asked that prior year tested status be included in the CAA/CIA data base to better track student requirements.

OSPI WASL Operations Director Christopher Hanczrik collected feedback about the well-received Webinar initiated last fall to take the place of the drive-in Regional Assessment Workshops cancelled this year due to the state budget crisis. A best assessment practices document is under construction by the Association of Test Publishers and the Council of Chief State School Officers. Feedback will be sought from test directors via a sub-committee. The resulting document will guide state department RFPs to vendors, Hanczrik said, and provide insights to industry practices. ANSI standards are a possible follow-on outcome. *The Washington Assessment Weekly* remains warmly received by directors and we learned from Robin Munson, student information systems director, that a monthly CSRS newsletter is under construction.

Assistant Superintendent for Assessment and Student Information Joe Willhoft provided a Q and A session for the gathered directors.

Q. What happened to the diagnostic tests promised for this school year?
A. While the legislature wished to implement, the I-Grant $2.5 million for test development has been frozen by the state budget crisis.

Q. Are translations of math and science tests into the six most common foreign languages still expected for 2009?
A. Likely just Spanish and Russian will survive plus a non-secure glossary of terms in English. Translation will be on a CD.

Q. When will we see the new grades 3 to 8 math assessments?
A. They will come online in 2010 to fit the new standards. Expect several versions of the pilot items in 2009 to fill gaps in the item banks. The pilots will be drawn from schools in the early return cycle.

Q. Will HS math have new standards?
A. Math continues with the same standards but new items are piloted for 2010 roll out.

Q. Will there be on-line assessments?
A. The current contract does not call for on-line testing but change orders (generally expensive) could be negotiated. This is year one of a 4-year development contract with ETS—renewable twice.
Q. Is an amended schedule possible for later in the year testing?
A. Not likely in 2009. We still have NCLB reporting deadlines.

Q. Will teachers score WASL items this year?
A. No, only COEs, a significant cost savings.

Q. May students bank COE scores?
A. No, they must be eligible for the option but those found to be eligible before the scoring window has closed may be scored.

Q. Why test WASL Math April 13/14, so early in the month?
A. While some schools are just back in session from spring break, the new contract has tight deadlines for score returns.

Q. How about the vertical scale for reading?
A. Reading grades 3 to 8 had four scaling methods. Two non-trivially different scales emerged. It is not clear which to adopt. Measurement expert Barbara Plake (U Nebraska) suggests that regression analysis may be as robust. New math standards have deferred scaling work in that content area.

Q. Was SPI-Elect Randy Dorn a Pearson employee?
A. No, he was an ETS and other testing company (non-measurement) contractor. He was a lobbyist for Pearson when the state contract was up for bid.

(Continued on next page…)

Q. Are there changes to the NCLB workbook?
A. No, the feds have given full approval. OSPI has not applied for the growth option. NCLB reauthorization is unlikely to proceed until 2011. Regulatory changes may occur.

Q. Why was WAAS-DAW moved to March for grades 11 and 12?
A. This happened last year so that scores could be returned on time, the day after Memorial Day.

Q. Why take the WASL when the SAT or ACT meets the requirement?
A. Neither national test is built to match Washington learning expectations. Cut scores for college bound seniors are reading-5%ile, writing-6%ile, and math 27%ile compared to the SAT.

Q. Do we know the WASL/WLPT correlation?
A. Yes, we have concordance tables. The exit standard on WLPT is similar to meeting the WASL standard.

-Editor's note: The world has shifted beneath our feet with the unveiling of Superintendent Dorn's plans for the state tests since the conference.

Willhoft Briefs Pre-Conference Assessment Audience

OSPI Asst. Supt. Joe Willhoft briefed an attentive pre-conference audience on the history and nuances leading to Full Approval of Washington's NCLB assessment system. The pre-conference workshop provided OSPI background on the two year struggle to move from "needs evidence" on all seven standards to the August 6, 2008 "Full Approval with Recommendations" letter. Further portfolio alignment is requested with continued teacher training.

Willhoft noted that the recent science alignment study spearheaded by Assessment and Psychometrics Director Yoonsun Lee looked closely at cognitive complexity, not simply item difficulty. Multi-step items have cognitive complexity, he explained, but uncommon knowledge (e.g. naming Neptune's moons) is simply difficult. A leveled analysis showed no Level 3 items:

Level 1 Remembering facts, concept comprehension…
Level 2 Application, analysis, synthesis, evaluation…
Level 3 Unclassifiable, ambiguous wording…

Test fit was found to be strong across all three grades (5, 8 and 10) with increasingly challenging content. The peer review was complementary, he said.

Math Assessments
High school mathematics expectations (119) appear in at least one of the three math course schemas (Algebra 1, Integrated 1, Geometry/Integrated 2, Algebra 2/ Integrated 3) plus a new HS WASL to reflect those standards. The end-of-course (EOC) exams may result in greater commonality across classrooms and districts, Willhoft said. There is no funding for Algebra 2 and Integrated 3. Some 26 of the 119 current targets would not be tested with those upper level assessments. Look for:

Pilot Items	Spring 2009
Pilot Items	Spring 2010
All five tests	Spring 2011

A comprehensive, common core test is envisioned for re-testers in the traditional WASL test window but the end of course tests would be later in the year. Middle school students would have to take both the WASL and EOC test to meet current AYP requirements.

The College Readiness Math Test (CMRT) developed through the University of Washington Office of Educational Assessment (OEA) with sign off by all public post-secondary institutions will be available on a contract basis to measure Algebra 2 proficiency. September 2009 is the go point, Willhoft said, but funding is uncertain. Those who meet standard would not be required to take remedial math classes at any of the Washington public universities or colleges.

(Continued on next page…)

The project website reports, "The new General Math Placement Test (MPT-G) has been distributed to testing centers at public four-year institutions, along with revised (three-option) versions of the Intermediate and Advanced tests (MPT-I and MPT-A, respectively). We are actively seeking administrative sites at two-year schools and the high schools to obtain data for setting the college readiness cutoff score." For a full accounting of the project see http://www.washington.edu/oea/services/testing_center/crmt/about_crmt.html.

WASL Gone?

A legislative committee will report on WASL reform options to the current legislation. SPI Randy Dorn has promised dramatic changes, Willhoft said: shorter turnaround for reports, fewer constructed response items, shorter administration time, and more information to parents and teachers. The budget is in critical shape, he noted. Willhoft assured the audience that the WASL will proceed as planned in spring 2009.

-Editor's Note: Assessment directors learned in early February that voluntary grade 9 testing was eliminated.

Praeger's Follies Winners Announced!

By Michael Power

The participants at the December WERA conference once again had the opportunity (some feel it's "the obligation") to compete for truly meaningful prizes in the annual Praeger's Follies event. Hosted by Bob Silverman and Michael Power, this event commemorates the many humorous contributions of long time WERA member Geoff Praeger who began this event many years ago and then got out of town.

The rules this year were simple, but the task was not. Using only the presentation titles in the WERA conference program, participants had to rearrange words or phrases to create a new title or sentence which was both comprehensible and clever enough to impress the contest chairs.

The winners and their winning entries were:

- Grand prize ($50 book store gift certificate) to Marty McCall of NWEA for her entry, "Social lunch fosters increasing middle for WERA faithful."
- Second prize ($25 book store gift certificate) to Annie Johnson of Mukilteo School District for her entry, "Students high on math may impact our investments."
- Two honorable mentions ($10 book store certificates) to Brian Rick of Bellingham School District for his entry, "On the Move—How to Score with Models."

Winners from Left: Marty McCall, Nancy Katims, Annie Johnson, and Brian Rick

and to Nancy Katims of Edmonds School District for her entry, "Social Strategies: The Dynamics of Embracing While On the Move."

Stay tuned for Praeger's Follies '09 at this December's conference.

Editor's Note

The Standard Deviation continues to grow as the association newsletter maintains long standing features and adds content customarily found in educational journals. The WERA Board of Directors has encouraged further development of the journal with the formation of an Editorial Advisory Board of members doing applied research in schools and the expansion of a volunteer editorial staff.

Initial members appointed to the Advisory Board are:

Janet Fawcett, Renton Schools
Janet Gordon, North Central ESD
Jill Hearne, Educational Consultant, WERA Past President
Yoonsun Lee, OSPI
James Leffler, Northwest Regional Educational Laboratory, WERA Board
Andrea Meld, OSPI, WERA Board Liaison
Brian Rick, Bellingham Schools
Prof. Michael Trevisan, Washington State University, WERA Past Board Member

The Advisory Board members have agreed to review the occasional submitted article and provide guidance about the direction of the journal. A fuller description of roles and responsibilities will be developed over the next several months.

The volunteer editorial staff are:
Peter Hendrickson, Editor, Everett Schools
Phil Dommes, Book Review Editor, North Thurston Schools
Don Schmitz, Photo Editor, Mukilteo Schools
Pat Cummings and Jack Monpas-Huber, Technical Co-Editors, Tacoma and Shoreline Schools
Michael Power and Bob Silverman, Humor Co-Editors, Tacoma and Puyallup Schools
Jeanne Willard, Editorial Assistant, Everett Schools

Websites of Interest to Measurement Folks: Program Evaluations *(continued)*

http://www.ed.gov/policy/gen/guid/fpco/hottopics/ht12-17-08.html
The Department of Education's "Dear Colleague" letter announces the publication of the new FERPA rules effective January 9, 2009. There are some new obstacles for program evaluators who are contractors.

http://www.wmich.edu/evalctr/jc/
The revised **Program Evaluation Standards (PgES) are ready for field trials and hearings**. Access to these sites' login details is available from The Joint Committee for Student Evaluation.

http://education.wsu.edu/aec/
The Assessment and Evaluation Center (AEC), located in the WSU College of Education, was established in 1997. The staff provides educational assessment and evaluation research and service to school districts, state agencies, university departments and other institutions. The work is supported through external grants and contracts.

WERA Board Members

Nancy Arnold, President
Director, Special Programs
Puyallup School District
214 W. Main
Puyallup, WA 98371
253-435-6532
253-841-8655 (fax)
arnoldnl@puyallup.k12.wa.us

Emilie Hard, President-Elect
Principal, Glacier Park Elementary School
Tahoma School District
23700 SE 280th Street
Maple Valley WA 98038
425-432-7294
425-432-6795 (fax)
ehard@tahoma.wednet.edu

Lorna Spear, Past President
Executive Director of Teaching and Learning,
Spokane School District

509-354-7339
509-354-5965 (fax)
lornas@spokaneschools.org

Leonard Winchell, Executive Secretary
Washington Educational Research Association
PO Box 64489
University Place, WA 98464
253-564-4816
253-564-4816 (fax)
lenwwa@aol.com

Members-at-Large

Phil Dommes
Director of Assessment
North Thurston School District
305 College Street NE
Lacey WA 98516
360-412-4465
360-412-4555 (fax)
pdommes@nthurston.k12.wa.us
(Term expires April 30, 2011)

James Leffler
Program Director, Services to the Field
Northwest Regional Educational Laboratory
101 SW Main Ste. 500
Portland OR 97204
800-547-6339 ext 549
503-275-9584 (fax)
lefflerj@nwrel.org
(Term expires April 30, 2009)

Andrea Meld
Research/Assessment Analyst OSPI
PO Box 47200
Olympia WA 98504
360-725-6438
360-725-6333 (fax)
Andrea.meld@k12.wa.us
(Term expires April 30, 2009)

Paul Stern
Senior Research Associate
Washington State University -SESRC
PO Box 43170
Olympia WA 98504
360-576-6030
360-586-2279 (fax)
sternpo@wsu.edu
(Term expires April 30, 2011)

The Educational Journal of the Washington Educational Research Association

Volume 1 ~ Number 1 ~ February 2009

Inside This Issue

On the Seesaw between State Assessment Work and University Teaching and Research

By Catherine Taylor, Ph.D.

In September, I returned to the University of Washington after a twenty-one month leave of absence during which I helped OSPI contend with federal No Child Left Behind assessment requirements and state graduation requirements.[1] This seesaw between OSPI and the University is not new for me.

Between 2003 and 2004, I also took a leave from UW and worked at OSPI as the Acting Director of Assessment until OSPI hired Joe Willhoft to serve in that role. Prior to that, I often spent 25-50% of my time on grants related to the state's assessment work. Colleagues within and outside of the University often ask why I am so involved in state assessment work.

Background

Prior to coming to Washington state in 1991, I worked for 10 years on state and national tests. As a graduate student (1980-1984), I worked on the Kansas Competency Testing Program. From 1984-1986, I worked for Harcourt, Brace, Jovanovich as a psychometrician on the Connecticut mastery testing program and the *Metropolitan Readiness Test*. From 1986-1991, I worked for CTB McGraw-Hill as a senior project manager responsible for seven test series including the *Curriculum Framework Assessment,* the *Early School Assessment,* the *Primary Test of Cognitive Skills,* and the *California Diagnostic Reading and Mathematics Tests*. When I left the private sector and came to the University of Washington, I was immediately recruited to help OSPI. Washington State had just adopted the *Curriculum Frameworks Assessment* as its 11th grade test and, because it wasn't a norm-referenced test, educators had no idea how to interpret scores. When Washington State's school reform law (HB1209) was passed in 1993, I was already viewed as a resource on testing.

The language of HB1209 established the Commission on Student Learning (CSL). The CSL, a panel of nine educational leaders, business leaders, and concerned citizens, was responsible for establishing essential academic learning requirements (EALRs) and a *performance-based assessment system* to determine whether students were achieving those learning requirements. Work began on the EALRs for reading, communication, writing and mathematics in 1994. Under the guidance of a national group of assessment experts (the National Technical Advisory Committee), the CSL decided that the assessment system would be composed of standardized tests in listening, reading, writing, and mathematics in grades 4, 7, and 10.

In the early 1990s, little was known about how to develop effective large-scale performance-based assessments. In fact, none of the major testing companies had successfully developed constructed response items in reading, mathematics, or science until 1990 – when Maryland contracted with CTB McGraw-Hill for the development of the Maryland School Performance Assessment Program.

In 1994, I received a grant from the CSL to develop and pilot prototype items and tasks in order to help the state select effective assessment formats. This was my first opportunity to engage with the state in conducting research. I worked with teachers from across the state as they wrote assessments ranging from multiple-choice items to multiple day projects. Staff at UW coordinated statewide pilots in 85 of the 296 school districts in the state. Washington's first assessment contractor (Riverside Publishing Company) presented these prototypes to teacher committees as they selected the types of items that would ultimately be incorporated in the Washington Assessment of Student Learning (WASL). Washington teachers selected, from among the prototypes, multiple-choice, short-answer, and extended response item formats for WASL.

Validity and Reliability Studies

By the late 1990s, I had become increasingly concerned about the validity of WASL scores. In 1998, I requested, and was granted, permission by the CSL to conduct validity and reliability studies for WASL and to present the results in WASL technical reports.[1]

(Continued on next page...)

To my delight and surprise, the studies showed that, although there were issues related to a few test items[1], there was strong evidence in support of the reliability and validity of overall WASL scores. Data showed that the scores were very reliable. In fact, score reliability was as strong for WASL as for more traditional standardized multiple-choice tests such as the *Comprehensive Tests of Basic Skills* (CTBS) and the *Iowa Test of Basic Skills* (ITBS). The results of studies provided very strong evidence that WASL tests truly measured what they were intended to measure. One surprise was that the results of the studies actually called into question the validity of ITBS and CTBS scores[1]. The studies suggested that the multiple-choice item formats for the ITBS and CTBS might be detracting from the validity of ITBS and CTBS scores.

Specialized Validity Studies

In 2000, I requested and was granted funding from the state to do additional validity studies to examine three technical questions related to WASL: 1) the stability of the scale scores over time, 2) how item types might affect the validity of scores, and 3) the effects of reading on WASL mathematics test scores.

Study 1: Examining the stability of scale scores over time is essential. Each year, new items are presented in WASL. This practice is done to prevent teachers from teaching specific test questions – a practice that was found to be widespread in 1980s and 1990s in states where test scores were used for high stakes decisions. Statistical techniques are routinely used to equate test scores from one test form to the next and the methods used in Washington State are state of the art. Still, the public must have confidence that the standards for passing remain the same despite the changes in items each year. The results of the study on the stability of scores showed that WASL scores were remarkably stable. That study (Taylor & Lee, in press) will soon be published in the journal, *Applied Measurement in Education*.

Study 2: WASL item analyses and the validity studies involving ITBS and CTBS raised our awareness of potential problems with multiple-choice test items. Issues of bias are a critical concern in testing. In our studies (e.g., Taylor & Lee 2008), we looked to see whether there was differential performance on items when comparing boys with girls and whites with Asian Americans, African Americans, Native Americans, and Latino Americans. What we found was staggering! When students were asked (on the reading WASL) to draw conclusions, make inferences, or interpret text (all of which are routinely assessed via multiple-choice items), multiple-choice items favored whites and constructed response items favored both whites and minorities; multiple-choice items favored boys and constructed response items favored girls. When students were asked to solve mathematical problems, reason mathematically, and represent mathematical ideas in graphic, symbolic, and other forms, multiple-choice items favored boys and constructed response items favored both boys and girls. These patterns were found across five testing years and at all grade levels; the patterns became more extreme as students got older.

Multiple-choice testing was first used in the early 1900s – long before anyone worried about equity or bias in education or testing. It was a simple, efficient, and 'objective' form of testing. Until the 1990s, with the exception of writing tests, very few large-scale tests included items that required students to construct their own responses.[1] Until the past few years, there has been little systematic data that would allow researchers to investigate bias due to item type. As we prepared the results of these studies for publication, we discovered that similar results had been found by researchers in other states and by the Educational Testing Service (ETS).

Study 3: The third study (Taylor & Lee, 2004) was designed to examine the influence of reading on WASL mathematics items. We conducted this study because of questions raised about the number of word problems on the WASL mathematics tests. We identified students who were proficient readers and struggling readers and compared their performances on WASL mathematics items using differential item functioning analysis (DIF). The results were surprising; the vast majority of flagged mathematics items favored struggling readers at all grade levels! Upon looking at the specific items that favored struggling readers, we found three item types: items that resembled textbook exercises, items that had graphic elements (e.g., bar or line graphs, geometric figures), and items that were 'story problems' related to issues relevant to students (e.g., using mathematics to select the best cell phone plan or to

(Continued on next page...)

determine which was more profitable – working on commission or having a straight salary). These results show that students who are struggling readers can do well with the more graphic aspects of mathematics and can work through contextual information when the problems are authentic to students. At present, we are preparing to publish this study. We have found no published research to date that examined this issue.[1]

The validity and reliability studies conducted for the WASL technical reports and the special validity studies described above, showed that WASL reading and mathematics tests are high quality tests. The scores are reliable and valid; the item analyses show that the items are technically sound; the balance of multiple-choice and constructed-response items added fairness to the test; the use of authentic problems allowed struggling students to demonstrate their mathematical knowledge and skills.

Mutual Benefits

It seems that, ever since I came to Washington, I have had one foot in the state's work and one foot in the University. I believe this has been a mutually beneficial relationship. As I did more research on WASL scores, I became more involved in helping to ensure the technical quality of the tests. I trained OSPI staff on how to critically evaluate the quality of items and how to interpret item analysis data. I was able to help OSPI deal with contractors because of my past experience as a test developer. In exchange, I have had access to significant amounts of data for research. I have been able to research ways to effectively develop performance-based items on large scale tests; I have been able to contribute to the increasingly large body of research about bias in multiple-choice testing; I have been able to investigate potential multi-dimensionality in mathematics testing. Most recently, through my work on the Washington Alternate Assessment System (WAAS) Portfolio, the Washington Language Proficiency Test, and the high school graduation assessment alternatives, I was able to use my classroom-based and large scale assessment expertise to help OSPI implement these new assessments, gaining both knowledge and data for future research in the process.

I came to UW to make a difference in the preparation of teachers – to empower them so that they could withstand pressures to teach to large scale tests. I have learned more from my students and the thousands of teachers I've met through my state work than they have learned from me. However, because of the state work, I have been able to bring facts, rather than fancies, to my teaching; I am better able to prepare my students for their future experiences with WASL and other state assessments. I have been able to help them understand the EALRs and, later, the Grade Level Expectations (GLEs), because I understand these targets better. Because a large focus of my teaching has been on how to develop valid and ethical classroom-based assessments, I have been able to help the state on the WAAS Portfolio and the Collections of Evidence.

In a better light, the seesaw I have been riding between state assessment work and university work is not a seesaw at all. The state work has improved my teaching and my research. When I mentor doctoral students, I always encourage them to consider working in a testing company or a school district before they go to academia: partly because they will have much better research questions and partly because they will be able to bring concrete, real world applications to their teaching.

References

Taylor, C. S. & Lee, Y. (in press). Stability of Rasch scales over time. *Applied Measurement in Education.*

Taylor, C.S. & Lee, Y. (2008, March). Gender DIF in Tests with Mixed Item Formats. Paper presented at the annual meeting of the National Council on Measurement in Education.

Taylor, C.S. & Lee, Y. (2004, December). Is Reading a Dimension in the WASL Mathematics Test? Using differential item functioning to examine the dimensions of WASL mathematics. Presented at the Washington State Assessment Conference, Seattle, WA.

– *Catherine Taylor, Associate Professor of Educational Psychology, University of Washington is a past WERA* president.

Shoreline to San Diego, an Accountability Tale

By Nina Salcedo Potter, Ph.D.

I should start by saying that I've always loved being on a college campus. Growing up in the San Francisco Bay Area, I always thought UC Berkeley was the coolest place for shopping and hanging out. I did not start out as one of the best students in college. I kept my grades above a C level so that I would not be put on academic probation, but I didn't have much more motivation than that. Then I took my first research design class in psychology. I was hooked. I loved statistics. It was the first time that I truly enjoyed math (well, actually, I loved doing proofs in geometry, but that is not something I usually admit to). Once I took that class, I thought, "Maybe I should try and get my GPA up, just in case I decide to go to graduate school."

I did not know at the time that I could actually major in research design and statistics. I did not even know there was something called psychometrics out there. But after a year of working at a daycare, I knew I had to go back to school. I loved working with little kids and had some experiences with kids with disabilities so I decided to get a M.Ed. in Early Childhood Special Education. As much as I loved working with the kids, it was once again the statistics course that I enjoyed the most. It was during my master's program that I learned that I could actually study statistics as my major. It didn't take long before I went back to school to get my Ph.D.

When I was finishing up my Ph.D. program I was unsure about the prospect of developing my own research agenda. While I loved using statistics and learning about statistics (and now psychometrics), I still did not have one area I saw myself focusing on. I thought that working at a school district in an assessment and/or program evaluation office would allow me to use my skills as a researcher in a variety of projects, and that a district position would be a better fit for me than working at a university trying to develop my own research agenda. I had taken part in a couple of program evaluation projects for school districts as a student and I really enjoyed that work.

As it turned out, the No Child Left Behind (NCLB) initiative made assessment become a much bigger part of the district assessment position and there was little time left for program evaluation. Tracking students, determining which students belonged to which underrepresented group and graduation requirements ended up taking up the majority of my time. When we had to hire a consultant to do a big program evaluation at one of the high schools because I did not have time to do it, I knew my time at the district was limited.

I had kept my eyes open for academic positions, but they were either in the Midwest, where I had no desire to move, or they were at Research Level I universities, and they wanted applicants with an active research agenda. I hadn't published anything since I graduated, and figured they would not have any interest in me. Then I saw a position at San Diego State University as the Director of Assessment for the College of Education. It was not a faculty position, but a management position. They wanted the same skill set that I was using in my job as Director of Assessment at Shoreline, but applied at the university level. Not only was a job at a university, it was in San Diego.

I've only been at SDSU for about six months, but it was the right move for me. I've had to learn about completely different assessment systems – Good-bye WASL, hello PACT (Performance Assessment of California Teachers) – and new educational systems, but it has been a positive experience overall.

Much of the work is very similar. SDSU is one of the 22 campuses of the California State University (CSU) system. I picture the central CSU administration kind of like OSPI and SDSU like a school district. The Dean of the College of Education (COE) is similar to the Superintendent of the school district. Each department within the college has a chair that serves much as a building principal, and each program within a department has a program coordinator which is much like the department chair of a high school or middle school. And of course, instructors are instructors at any level. Whereas in P-12 there were grade level expectations (GLEs) to guide what is taught at each grade level, we have state and national standards that we have to ensure every student learns.

(Continued on next page…)

The legislature in California passed a law stating that in order to receive a teaching certificate, candidates (students in the teaching program) had to pass an assessment (PACT is one of two options) just as the Washington legislature passed a law that in order to receive a high school diploma all students had to pass the WASL. There are meetings with representatives from all the CSUs to discuss how we will implement the assessment just as all of the Assessment Directors in WA met to discuss how we were implementing WASL.

At this point, there is only one state mandated test at the university and it is only mandated for initial teaching credentials. None of the other programs (i.e. Educational Leadership, School Counseling, School Psychology, Special Education, and Educational Technology) have state mandated tests. Of course the trend does indicate this could change. A big difference between the PACT and WASL is that we have to score the PACT ourselves. With over 500 students finishing every year and two to four hours to score each assessment, coupled with no funding from the state, you can imagine the conversations surrounding this. I don't think I have to say much more about that!

Much like at the school district, I work with people across all levels collecting and analyzing data to help improve courses, programs and the COE as a whole. Part of my job is to help each program determine what assessment data they will collect to evaluate their program. Much like a school district there are some programs way ahead of the curve in terms of data collection and analyzing and some further behind.

A big difference for me is the amount of help I receive and how my time is spent. I have a full-time research assistant who does most of the work with the database and creating tables and graphs. I can choose which project I want to spend my time on. There is a full-time data administrator who, among many other responsibilities, works with the student information system and other data so that my expertise can just be assessment data.

My time is really spent working with faculty to determine what assessment data is best for evaluating their programs, determining which assessment data we should include in the assessment system of the college as a whole, and other work related to the COE assessment system. There are times I feel that I spend more time in meetings than I'd like, but at least I have enough support that the work still gets done.

Another completely new thing for me is learning about accreditation. Any program that ends with a credential has to be accredited by the state and the COE as a whole is accredited by NACTE at the national level. I am the coordinator for both of these processes and learning about the whole accreditation process has taken up most of my time. I have been happy to discover that the movement in accreditation is towards a more data driven process. We are required to create an assessment system that includes student outcome assessment data. Not all faculty members are on board with the idea that they should be required to do the same assignment or assessment as all other faculty teaching the same course. But seeing as this is a college of education, I think this move is easier than it is at other colleges. Since our faculty teach about the importance of using student level data in education, it's hard for them to argue that they should not be doing it themselves. I know from discussions I've had with colleagues from other colleges, this is a much more difficult movement in other colleges such as business or engineering.

As I mentioned before, this position is not a faculty position. I am not tenure track and therefore I do not have any requirements in terms of publishing or teaching. However, I do love research and I am trying to build relationships with faculty members so I can help with their research. I may also begin serving on Dissertation Committees, especially for those dissertations involving quantitative research. I am also hoping to begin teaching a course next fall.

I have learned that my position is rare. COEs often have assessment directors or coordinators, but the job is typically given to a faculty member in place of some of their teaching and/or research requirements. They still have to be sure to get published if they want to get tenured.

Colleges or departments outside of education rarely even have assessment coordinators. This is changing however as there is more and more push for outcomes assessment in higher education.

(Continued on next page…)

Every department at SDSU has to do an annual outcomes assessment report. They are required to develop learning goals and objectives, develop direct assessments of student learning to measure these goals and then to use the results to make programmatic changes. I serve on the committee that heads this and there is a lot of work to be done in this area.

-Nina Salcedo Potter, Director or Assessment for the SDSU College of Education, was Director of Assessment in Shoreline Schools and active in WERA.

The Role of the Statewide End-of-Course Assessments in High School Assessment Systems: A Study for the Washington State Board of Education

By Sionain Marcoux

For more than 15 years, Washington policymakers have sought to continually improve public K-12 education so that all children are expected and taught to learn at high levels. The Washington State Board of Education (SBE) recently set ambitious goals for its oversight of the K-12 public education system: "Raise student achievement dramatically" and "Provide all students the opportunity to succeed in postsecondary education, the 21st century world of work and citizenship" (SBE, 2006).

In 2007, the Legislature enacted ESSB 6023, which directed the SBE to examine and recommend changes to high school assessments with a limited series of end-of-course (EOC) assessments. Governor Gregoire vetoed this provision because she felt the study should not predetermine that end-of-course assessments would be implemented. Instead, she asked the SBE to study policy and technical issues about EOC assessments.

To inform the deliberations of the Governor, SBE, OSPI, legislators and interested stakeholders, the SBE contracted with Education First Consulting, LLC, to conduct an independent study of statewide end-of-course assessments. This report summarizes the findings of our research study across several lines of inquiry:

- What lessons can Washington state learn from the literature on high school assessment and accountability systems, with a focus on EOC assessments and high school exit exams?
- What have been the experiences of other states in implementing EOC assessments?
- Do other assessments measure the same content and skills as the WASL?
- What are the policy implications for Washington's high school assessment system, based on the literature and lessons learned from other states?

To address these questions, Education First Consulting conducted a thorough review of the primary and secondary literature on EOC assessments and high school assessment and accountability in general. To develop a picture of the diverse ways EOC assessments are deployed across the nation, we reviewed EOC programs in nine states—California, Indiana, Maryland, New Jersey, New York, South Carolina, Tennessee, Texas and Virginia. After this initial environmental scan, we conducted 30 interviews with key education, government and business leaders in six states—California, Indiana, New Jersey, Tennessee, Texas and Virginia—to obtain more in-depth knowledge of states' experiences with EOC testing.

Defining Comprehensive and End-of-Course Assessments

We define comprehensive assessments (also known as end-of-grade tests) as measures that assess a range of material in a particular subject area. The material may have been taught in previous grades and via different courses, but this common test is administered to all students in the same grade near the end of the school year. Most states administer comprehensive assessments just once in high school, typically in grades 10 or 11, and all eligible students in that grade take the test. While it is most common for states to administer comprehensive assessments in language arts and mathematics, many states offer comprehensive assessments in the four core academic subjects—language arts, mathematics, social studies and science.

We define end-of-course tests as assessments designed to measure mastery of standards for particular high school courses. EOC assessments are administered on a more flexible schedule, since the tests are administered only to those students who take the course. The major reasons states cite for giving EOC assessments are to assess learning of specific course content and to administer the tests closer to the time of instruction. Unlike

(Continued on next page…)

comprehensive assessments, which measure content areas, such as mathematics, EOC assessments are designed to correspond with learning standards in specific courses, such as Algebra I, English II, U.S. History or Biology. EOC systems hold the course the student takes, not the grade level of the student, constant. For example, in the most extreme cases, students in middle school and in 12th grade may be included in assessments for Algebra I.

Key Findings and Lessons Learned

High school assessment systems have four major purposes.

High-quality high school assessment systems are an important tool for:

- Supporting student learning by measuring achievement of state academic standards and diagnosing academic strengths and weaknesses
- Holding students and/or schools accountable
- Determining readiness for postsecondary education and training
- Ensuring high-quality and efficient operations is a fourth major goal so that assessments produce sufficient information to meet the first three purposes well, while minimizing the costs and time spent on testing

Comprehensive and end-of-course assessments have different strengths.

This research shows that standards-based comprehensive assessments and standards-based end-of-course (EOC) assessments, on balance, can serve the four major purposes equally well. For example, both formats can diagnose student academic strengths and weaknesses; both formats are used as high school exit exams; and both formats can place students into credit-bearing college classes. But this report also shows that the formats have different strengths and meet these major purposes in distinctly different ways.

State high school assessment systems that are built around comprehensive tests:

- Usually focus on 10th grade or lower standards
- Assess a slice of the high school standards, rather than deep knowledge of subjects
- Can potentially narrow the delivered curriculum to what is tested
- Provide a "snapshot" of system performance at a common point in time for all students
- Often take up less testing time overall and cost less
- Take a more straightforward approach to exit exams and school accountability
- Rarely provide information on students' readiness for postsecondary education coursework and training

State high school assessments systems built around end-of-course testing:

- Vary widely with respect to the number and kinds of courses that are assessed
- Will measure a broader and deeper range of standards, including advanced subject matter, but only if there are a sufficient number of EOC assessments in each subject
- Do not assess all students against common standards unless states require all students to take a certain series of courses and/or require all students to take certain EOC assessments
- Are typically implemented to promote more consistency of teaching and provide more timely information on learning and course quality
- Motivate students to learn through exit exams as well as other forms of lesser student stakes, such as counting test results as a portion of course grades
- Make it more complicated to hold students and schools accountable, yet offer the potential to produce more validity and reliability
- Can be better suited for placing students in postsecondary education courses than comprehensive tests given by states in the 10th grade

We also learned that changing test formats does not necessarily improve student learning of state standards or increase student performance. And states are now permitted to use EOC assessments to meet the requirements of No Child Left Behind. Finally, other studies have shown that alternative

(Continued on next page…)

assessments to the WASL vary in the degree to which they measure the full range of skills and knowledge found in the WASL.

Policy Implications for Washington

Table 1 below shows clearly that, while the two formats can serve many similar purposes, they also have different strengths in different areas. Given that comprehensive and EOC assessments have much in common, and that neither format is in itself a panacea to problems of low student or school performance, Washington policymakers must first determine the extent to which the four purposes are most important in Washington, in order to choose the most

Table 1

How Well Do Comprehensive and EOC Assessments Meet the Four Major Purposes of Assessments?

Issue Area	Advantage to…
(1) Supporting Student Learning	
Measuring the breadth and depth of standards	EOC assessments (slight)
Assessing students near the point of curriculum delivery	EOC assessments (strong)
Assessing students with the same test	Comprehensive
Choice and quality of test question types	No clear advantage
(2) Holding Students and/or Schools Accountable	
Validity and reliability of assessments	EOC assessments (slight)
Holding students accountable	No clear advantage
Reporting results at the classroom or course level	EOC assessments
Holding schools accountable	No clear advantage
(3) Determining Readiness for Postsecondary Education	
Measuring readiness for postsecondary education	EOC assessments (strong)
Providing access to rigorous courses while preserving flexibility	EOC assessments (slight)
(4) Ensuring Quality and Efficient Operations	
Testing window and turnaround time for results	No clear advantage
Costs and time spent on testing	Comprehensive
Impact of administration on schools	No clear advantage
Test security	No clear advantage

Outcomes

The research provided the governor, Legislature, state board of education and the Office of Superintendent of Public Instruction with the information needed to determine the most appropriate testing methods for Washington. In 2008, the Legislature changed the 10th-grade test in mathematics to two end-of-course tests in Algebra I and Geometry. While there are strengths and limitations to any assessment system, this research enabled the state to adopt substantially more informed policy.

–Sionainn Marcoux, Education First Consulting, presented to the December State Assessment Conference with Bethany Gross, University of Washington.

Summary of the New OFM Education Research and Data Center

By Deb Came, Ph.D., and Carol Jenner, Ph.D.

RCW 43.41.400: established an "education data center" in the State of Washington's Office of Financial Management. Jointly with the Legislative Evaluation and Accountability Program (LEAP) committee, the Education Research and Data Center (ERDC) is directed to:

- Conduct collaborative analyses of early learning, K-12, and higher education programs across the P-20 sectors
- Compile and analyze education data, disaggregated by demographics
- Collaborate with LEAP and legislative committees to identify data to be analyzed to ensure legislative interests are served
- Track enrollment and outcomes through the Public Higher Education Enrollment System (PCHEES)
- Assist in developing long-range enrollment plan for higher education
- Provide research that focuses on student transitions in early learning, K-12, and postsecondary education
- Make data available to agencies that contribute to ERDC, to the extent allowed

ERDC has partnerships with numerous agencies, including Department of Early Learning, Office of Superintendent of Public Instruction, State Board for Community and Technical Colleges, Higher Education Coordinating Board (HECB), Public baccalaureate institutions, Professional Educators Standards Board, State Board of Education, Employment Security Department, Department of Social and Health Services, and the Workforce Training and Education Coordinating Board.

In the first year and a half of the data center, ERDC staff have:

- Created a source for commonly asked questions, education indicators and standard education information: www.erdc.wa.gov
- Developed a preliminary longitudinal data system spanning K-12 and public higher education. Also, ERDC now has the capability to link with wage records and public assistance data.
- Added a longitudinal component to the Public Centralized Higher Education Enrollment System
- Established data-sharing agreements with partner agencies
- Developed data-linking processes and an anonymization protocol
- Contributed analysis to the legislatively mandated Per-student Funding Study (RCW 28B.15.068)
- Participated in K-12 data feasibility study and HECB strategic planning

ERDC has the capability to answer numerous research questions, but will focus on state-level analysis (with the possibility of drilling down to a more local level) and transitions data. A few examples of potential research questions are:

- What are the outcomes for those who drop out of high school? How many re-enter high school, get a GED, enter the workforce, or enroll in postsecondary education or training?
- What degrees and majors are pursued by students entering from high school? Are they different than students who enter as community college transfers?
- Were students who received need-based financial aid in college classified as eligible for FRPL in high school?
- To what extent do high school and college students participate in the workforce?

- Deb Came and Carol Jenner, Education Research and Data Center, Forecasting Division, Washington State Office of Financial Management.

WAAS vs. Transition Planning – Goal Writing and Instructional Challenges for Secondary Special Education Teachers

By Sara Woolverton, Ph.D. and Peggy Thesing

Ongoing changes in WAAS portfolio requirements (driven by NCLB) paired with the continuing clarification of the transition planning component of the Individualized Educational Program or IEP (driven by IDEiA 2004) present secondary special education teachers with a heightened challenge when drafting IEP goals and planning instruction for the state's most academically challenged students (OSPI, 2008).

Students with severe or profound cognitive disabilities are usually educated in programs geared toward instruction in independent living skills that become increasingly functional in nature (rather than traditionally academic) as students rise through the grades. For example, a middle or high school student without basic reading skills will have IEP reading goals and instruction emphasizing recognition and understanding of safety and function symbols (e.g. a stop sign or restroom symbol).

Because paper and pencil tests are not meaningful for these students, evidence of meeting standards is presented to the state via the Washington Alternative Assessment System (WAAS) Portfolio. The WAAS-Portfolio is an alternate assessment of a student's knowledge and skills based on evidence of student work that demonstrates progress over time and generalization of skills in various contexts. Students are assessed on the same Grade Level Expectations (GLEs) as their peers but expectations are adjusted to match the achievement levels and learning characteristics of each individual student (Kraft, 2008).

WAAS rules specify that goals and instruction for these students be linked to the targeted skills described in GLE extensions. Thus the teacher of the significantly cognitively delayed student with the symbol reading goals mentioned above might be expected to instruct toward Reading GLE extension. 2.1.5 HS.C which states, "The student will identify an inference/prediction and support it with two or more details from grade level text" (OSPI, 2008).

While the state continues to refine WAAS requirements IDEiA 2004 changes are driving further clarification of proper IEP transition plan formulation. OSPI has clarified that the IEP of each student aged 15 and older must contain post-secondary outcome statements regarding education and training, employment, and as necessary, independent living. The IEP team must write a descriptive statement identifying a specific post-school education/training activity (e.g. "will attend a vocational training program to learn culinary arts skills") and a specific employment goal ("will be employed in the food preparation industry") and for most life skills students, an independent living goal ("will live in a group home and use public transportation independently") (OSPI, 2008).

Transition planning for secondary students is intended to be the first part and cornerstone of IEP development. The bulk of the IEP guides instruction that will provide students with skills needed to meet the post-secondary goals identified in the transition plan. In this framework, students' academic IEP goals must describe functional skills that provide a foundation for the specific post-secondary education, training, employment, and independent living outcomes identified in the transition plan.

Many (if not most) secondary special education teachers experience the WAAS and transition planning demands as contradictory and often mutually exclusive. When teachers are expected to teach toward both grade-appropriate targeted skills and goals focusing on developmentally appropriate functional life skills they essentially feel forced to employ dual curricula toward competing outcomes in response to disparate regulatory demands.

The burdensome nature of the WAAS portfolio has been an issue for special education teachers since its inception. Increasingly special education administrators are hearing from the teaching corps that WAAS portfolio requirements are also leading to a departure from functional instruction. All secondary life skills teachers in this district and those who share our Educational Service District portfolio trainings have expressed concern over this issue.

(Continued on next page…)

As one teacher expressed it, the best spin she can put on teaching to GLE extensions is that they provide "enrichment activities" that may engage some of her students. The plea of our special education teachers is to be allowed to refocus instruction on skills that are useful and functional and will maximize the likelihood that our cognitively challenged graduates can live independent and fulfilling lives.

References

Kraft, Judy, (2008). *Washington Alternate Assessment System Teacher's Guide to the WAAS-Portfolio*, prepared by Judy Kraft, Alternate Assessment Specialist. Retrieved January 10, 2009 from the OSPI Web site: http://www.k12.wa.us/assessment/AlternativeAssessment/pubdocs/TeachersGuidetotheWAASPort.pdf

Office of Superintendent of Public Instruction, (2008). *Special Education Evaluation and IEP Technical Assistance Module*. Retrieved January 16, 2009 from the OSPI Web site: http://www.k12.wa.us/SpecialEd/Module.aspx

Office of Superintendent of Public Instruction, (2008). *Washington Alternate Assessment System (WAAS) Portfolio: Parent Guide 2008-2009*, Parent information pamphlet. Retrieved January 10, 2009 from the OSPI Web site: http://www.k12.wa.us/assessment/pubdocs/PortfolioBrochure.pdf

Office of Superintendent of Public Instruction, (2008). *Washington Alternate Assessment System: Extensions to the Grade level Expectations (GLEs) Reading*. Retrieved January 10, 2009 from the OSPI Web site: http://www.k12.wa.us/assessment/AlternativeAssessment/pubdocs/ReadingExtensions.pdf

-Sara Woolverton, Special Education Area Director and Peggy Thesing, Special Education Facilitator, Everett Public Schools, work closely with assessment staff on measurement issues.

The Care and Feeding of Doctoral Students

By Peter Hendrickson, Ph.D.

Urban school districts of a certain size with access to local graduate schools tend to generate doctoral students from within the staff. So often they are among the most capable teachers and administrators who are already fully engaged with their school careers and life-at-large. Doctoral students may feel their program is ill understood by colleagues, even by their supervisors who have not been through the committee and dissertation fires.

At present six teachers and administrators in our district of nearly 19,000 students are known to be in doctoral programs. They represent the University of Washington, Seattle University, Washington State University, and the University of Oregon. The programs range from educational leadership to second language acquisition. Some are in their first year and one is defending her dissertation in a few days. My own studies were completed over 20 years ago at the University of Washington.

It appeared that while each of the students had an advisor and many had a committee, they did not necessarily have a support group for the rigorous journey. Names were gathered from administrative colleagues and we first met for an hour after school in my office two years ago. The format was simple—tell your story, round robin, and let the conversation follow. I supplied one box of cookies and bottled water. My facilitation required only the lightest of touches as they were most interested in each other's work. Those who were a cohort ahead of a colleague told them what to expect and provided living proof that you really could make it to the next step while working full time with three kids at home.

We decided to plan future meetings after work hours (if high school principals have such a time) but off campus at a local brew pub. Meetings there are also scheduled for an hour but as the personal relations have flourished, the conversations have lengthened. The students now buy their own beverages and several extend the hour to enjoy the pub food. We meet five or six times a year and bear some relationship to a Professional Learning Community (DuFour et al, 2004). In this case the learning is distributed across institutions and the topics are diverse. The common threads are a collective thirst for scholarly work within the fabric of public schools and parallel journeys into the terra incognita of comps, committees, dissertations and capstone projects.

My professional reading yields resources which they may not have encountered in their studies such as the AERA Standards for Reporting Empirical Social Science Research in AERA Publications (2006). Early each year I provide copies of the district's educational research protocols and I make sure they're current with the most recent issue of *The Standard Deviation*, the Washington Educational Research Association newsletter/journal which I edit. I'm always scouting for journal authors and have mined the doctoral group for articles as I've encouraged them to present at regional conferences.

While the students benefit from each other's experiences and research areas, I've benefited, too. Colleagues have become friends, even co-investigators. There are no surprises when one wants to conduct a study as we have talked long before permissions are sought. I've come to know some of their advisors and have a better sense of current doctoral programs. And each of them has suggested an article or two I really should be reading.

This no-cost, minimal preparation activity helps establish our status as a community of learners. I look forward to relationships lasting long after they've defended and basked in the glow of congratulations from colleagues, spouses and children.

(Continued on next page...)

References

American Educational Research Association. (2006). Standards for reporting on empirical social science research in AERA publications. [Electronic version]. *Educational Researcher*, 35, 6, 33-40.

DuFour, R, DuFour, R, Eaker, R. & Karhanek, G. (2004). *Whatever it takes: How professional learning communities respond when kids don't learn*. Bloomington, IN: National Educational Service.

Hendrickson, P. (Ed.). (2008). *The Standard Deviation*. Accessed November 17, 2008 via http://www.wera-web.org/pages/publications.php

-Peter Hendrickson , Everett Assessment Specialist, is a Past President of NATD and WERA. Reprinted with permission from the NATD 2008 Newsletter.

Book Review: Editor's Note

Larry Ainsworth will be one of the keynote speakers at the Spring WERA conference. His book, Common Formative Assessment, was reviewed in the fall *Standard Deviation*. You can access this review at http://www.wera-web.org/links/TheStandard%20100608.pdf.

Book Review: *Outliers: The Story of Success* by Malcolm Gladwell

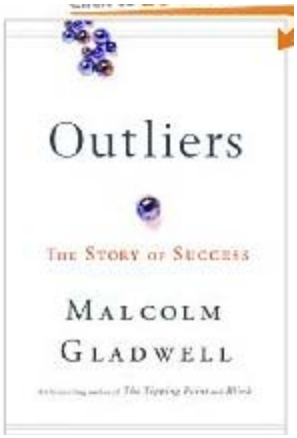
Reviewed by Lorna Spear, Ed.D.

We've all heard the stories of highly successful people and how their success hinged on their intelligence and desire to be successful. What if, instead, individual success is determined not by "extraordinary talent but…extraordinary opportunities?" Gladwell illustrates his thesis by sharing the backgrounds, lives and opportunities of individuals as diverse as Marita (a fifth grader in inner city New York City), Laureano Caviedes, (pilot of doomed Avianca flight 052 that crashed on approach to New York City's Kennedy Airport in 1990), and Bill Gates. While examining the lives of outliers or people with extraordinary accomplishments, the reader begins to see patterns of opportunities that have led to these achievements. His intriguing stories might lead you to reflect on your own career path, where you find yourself on the road to success, and even how you define success.

Outliers is a quick, thought-provoking 285 page read. His book left me wondering if we, in education, could do what Gladwell suggests on page 268, "To build a better world we need to replace the patchwork of lucky breaks and arbitrary advantages that today determine success—the fortunate birth dates and the happy accidents of history—with a society that provides opportunities for all."

Publication data: *Outliers: The Story of Success* by Malcolm Gladwell, Little Brown & Company, New York City, NY, 309 pages, 27.99, (US) ISBN 978-0-3160-7923

-Lorna Spear is an Executive Director for Teaching and Learning Services with the Spokane School District and is a current board member and past president of WERA.



Book Review: *A Whole New Mind –Why Right Brainers Will Rule the Future* by Daniel Pink

Reviewed by Monica Sweet

As a business and technology writer, Daniel Pink is well aware of the attributes that tomorrow's workers will need to ensure success in their chosen fields. In his latest book, A *Whole New Mind: Why Right-Brainers Will Rule the Future,* Pink convincingly argues that the days of the "information age" are behind us and that the new "conceptual age" has emerged.

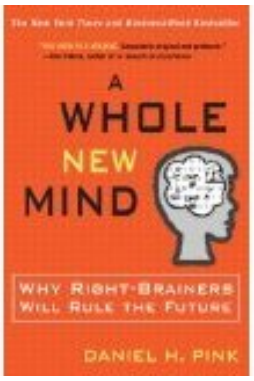
While the book focuses on the professional and personal lives of adults, Pink argues that this simple workforce shift from analysis and logic to empathy and creativity will have a profound effect on the world of instruction and students.
The first part of the book takes the reader through a short review of left and right brain functions in order to illustrate that "the future belongs to a different kind of person." Pink defines this person as someone who has right brain aptitudes (design, harmony, story-telling, empathy) in addition to left-brain thinking (analysis, logic, synthesis). While the study of the brain hemispheres may not be new, the impact of the research for today's students is quite significant. For example, Pink shares that business schools are slowly starting to recognize the power of narrative in the workplace. Successful companies are now seeking individuals who can not only create and sell a product, but also possess the ability to design a compelling vision and deliver it with emotional impact.

The second part of the book focuses on what Pink defines as the six essential right-directed aptitudes. He describes that these six senses, when coupled with left-directed reasoning, "…can help develop the whole new mind this new era demands." A chapter is devoted to each of the six senses: Design, Story, Symphony, Empathy, Play, and Meaning. Pink ends with a portfolio section that offers exercises and additional reading suggestions to actually put into practice the ideas presented.

A Whole New Mind is an outstanding read that not only reinforces the importance of a well-rounded education inclusive of right-brained coursework and opportunities, but also serves as a practical guide to ensure success in the personal and professional life of the reader. Teachers, principals, and students would greatly benefit from Pink's ability to share more than just theory and rhetoric; he is a storyteller who provides specific, detailed activities that allow all of us to make the best use of both sides of our brain.

Publication Data: *A Whole New Mind: Why right-Brainers Will Rule the Future* by Daniel Pink, 2006. Riverhead Books, New York, NY, Paperback, 275 pages, $15.00 (US) ISBN: 9781594481710

-Monica Sweet is the Principal of Chinook Middle School for the North Thurston Public Schools in Lacey.



Editor's note: Aspiring book reviewers are invited to contact Book Review Editor Phil Dommes, North Thurston Schools, at pdommes@nthurston.k12.wa.us

U W Brain Injury Study Uses WASL Percentiles

By Nancy R. Temkin, Ph.D. and Jin Wang

Traumatic brain injury is one of the most devastating injuries of childhood. Each year in the US over 6000 school-age children die from their injuries, 47,000 are hospitalized, and 364,000 visit an emergency department. Little is known about the more detailed functional deficits and course of recovery of the survivors. Dr. Fred Rivara of University of Washington and Seattle Children's and his team are trying to fill some of that gap. They are conducting a population-based study of children who get a TBI in King County, WA. Children hospitalized or coming to an emergency department with TBI and comparison children with other injuries are enrolled soon after injury and are followed at three months and one, two, and three years after injury to see how they are doing. This is measured in a variety of ways. For school-age children, an important outcome is academic achievement. We want to know if children with TBI, especially the most common but least studied mild TBI, lose ground or whether they stay up where they were prior to their injury.

Unfortunately, the study does not have the resources to bring every child in for an assessment that includes academic performance. The WASL provides a reliable standardized test taken by almost all children in Washington between 3rd and 10th grades. Its usual criterion-based outcomes are not ideal for answering the question of interest. They are quite coarse grained, and if, as everyone hopes, the pass rate is improving, staying in the same category might actually mean falling behind uninjured peers. Although raw scores could be used as the outcome, raw scores differ between subjects and across grades. We felt that using percentiles, that is the percent of students statewide who scored no better on the same exam than a student in the study, would allow us to fairly account for changes in the difficulty of the test or in the preparation of the students. We are collecting test performance for three years before the injury and up to three years after injury so we can sharpen our comparison by accounting for how a student had been doing before they were injured. We were surprised that percentiles were not readily available for the WASL. The OSPI Assessment Office kindly provided the raw data to us. In case others might have a use for the percentiles, we provide them here. Link: Tables

Tables 1 to 3 present the raw score that corresponds to each percentile for each test and grade for 2006, 2007, and 2008. A score is at the p^{th} percentile if at least p/100 of the students in that grade who had a valid score on that test in that year received that score or lower and (100-p)/100 received that score or higher. For example, in 2008, what was the 50^{th} percentile or median score for 4^{th} graders in math? Looking in Table 3 in the row labeled 50 and the column labeled Gr 4Math08, we see that the median score was 403. Let's check that a score of 403 satisfies the definition for the 50^{th} percentile.

Looking in Appendix spreadsheet percentilesmath2008 in the tab for Gr4, the scores are shown in column B and the number of students getting that score is in column C. The total number of students getting any valid score is in the last row of column C (74,785). For 403 to be the 50^{th} percentile, at least 74785*50/100=37392.5 students should have scored 403 or lower and 74785*(100-50)/100=37392.5 should have scored 403 or higher. Column E gives the number getting that score or lower. We see in the row labeled 403 that 39462 students scored 403 or lower. Column G gives the number getting this score or higher. For a score of 403, 37783 scored 403 or higher. Since both of these numbers are at least as high as needed, this confirms that 403 is indeed the 50^{th} percentile.

Not every percentile has a different score. In fact, on the reading test where there are few different scores observed, one score can cover many different percentiles. For the TBI study, we have the student's scores and we want to get the percentile that best corresponds to that score. To do that, we act as if the scores were continuous and used the standard convention for histograms of grouped continuous

(Continued on next page)...

data. That is, we acted as if half the students who got a particular score actually scored a little below and half a little above. Thus, for a score of 400 in the test above, we have 34687 scoring 399 or less and 2315 scoring 400, so we assign a percentile of 100*(34687 +2315/2)/74785 = 47.9% or round to 48%. This is one of the percentiles associated with a score of 400, but now we have a particular one to use. The percentiles corresponding to each score are given in appendix spreadsheets a-q in column D.

Since the number of students taking each test is so large, we didn't have to deal with some details such as 2 values satisfying the definition for the same percentile. For example, if there were 10 students taking a test and one scored each value from 1 to 10, the 30th percentile needs at least 10*30/100=3 students scoring that or less and at least 10*(100-30)=7 scoring that value or more. There are 3 students scoring 3 or less and 8 scoring 3 or more and 4 students scoring 4 or less and 7 scoring 4 or more, so both 3 and 4 meet the definition of the 30th percentile. The usual convention is to call the value halfway between the ones that qualify as the percentile. So the 30th percentile would be a score of 3.5. This sometimes means a value listed as a percentile is a value you could never actually observe. If the sample size is large, this rarely occurs, and never did in calculating the WASL percentiles.

When might you want to use percentiles, when proficiency cutoffs? All the students can be proficient, but the only place they are all above average is Lake Wobegon. Julie Hoff at OSPI reports that standard setting for the reading and mathematics assessments at grades 3, 5, 6 and 8 took place in 2006 following the operation Spring 2006 test. The standards for grades 4, 7 and 10 were revised in 2004. Grade 4 standards were originally set in 1997, grade 7 in 1998 and grade 10 in 1999, all following their initial operational administration.

Appendices

Raw score to percentile tables

-Math 2006 percentiles
-Math 2007 percentiles
-Math 2008 percentiles
-Reading 2006 percentiles
-Reading 2007 percentiles
-Reading 2008 percentiles
-Writing 2006 percentiles
-Writing 2007 percentiles
-Writing 2008 percentiles

-Nancy Temkin, is a Professor of Biostatistics at the University of Washington whose primary research area is traumatic brain injury.
-Jin Wang is a Research Consultant to Surgery at the University of Washington.

Stupid Excel Tricks for Assessment Folks

By Patrick Cummings

Introduction

The following is a simple Excel trick that converts text that is all capitals (like **SMITH**) to a proper format (like **Smith**). Many times we deal with long lists where the text in the data file is in all upper case letters. Converting text to different formats can be a useful tip when handling large data sets.

Text Syntax Overview

Let's start with some basic text syntax formats using the name **Smith**:

	A	B	C
1	**Original Text**	**Syntax Formula**	**Changed Text**
2	smith	=**UPPER**(A2)	SMITH
3			
4	SMITH	=**LOWER**(A2)	smith
5			
6	SMITH	=**PROPER**(A2)	Smith
7			
8			

UPPER changes text to all capital letters

LOWER changes text to all lower case letters

PROPER capitalizes the first letter and make all other letters lower case

The Irish Problem

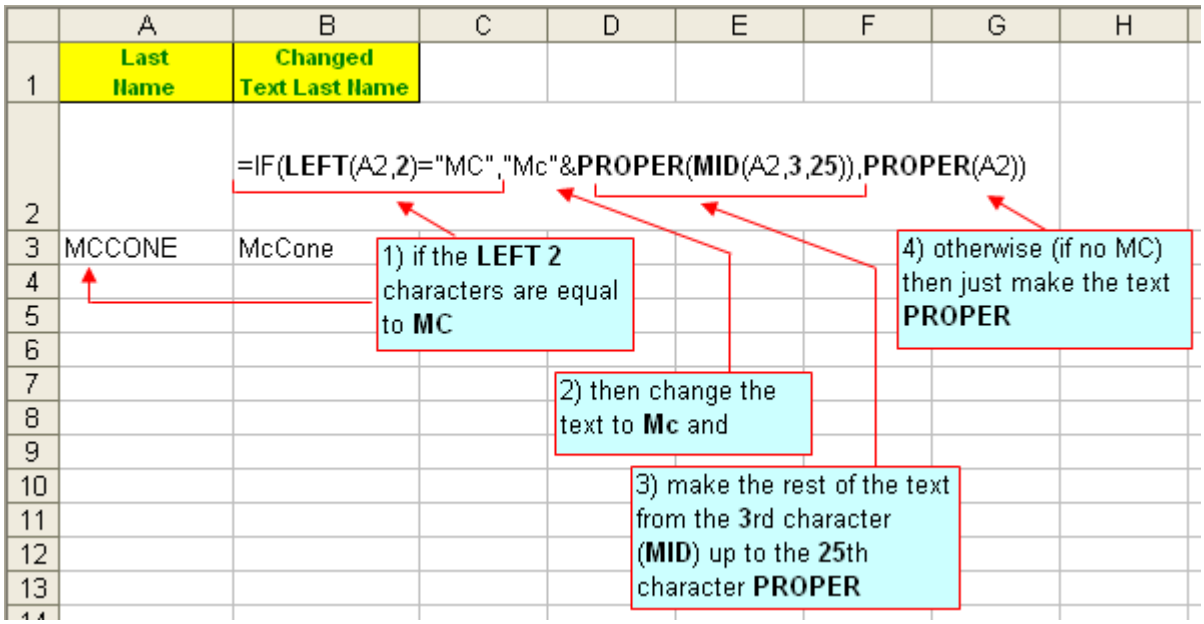
If we have a long list of names that are all upper case then the **PROPER** formula works fine until we get to those pesky Irish names like **Mc Cone**, **Mc Donald**, **McGee**, etc:

	A	B	C	D
1	**First Name**	**Last Name**	**Changed Text First Name**	**Changed Text Last Name**
2			=**PROPER**(A2)	=**PROPER**(B2)
3	ROB	JUDSON	Rob	Judson
4	MARGARET	KINSEY	Margaret	Kinsey
5	GREGG	LEACH	Gregg	Leach
6	ROBERT	LIGHTFOOT	Robert	Lightfoot
7	TONY	LUND	Tony	Lund
8	KATHY	MARTINEZ	Kathy	Martinez
9	LINDA	MCCONE	Linda	Mccone
10	KAREN	MCDONALD	Karen	Mcdonald
11	STEPHANIE	MCGEE	Stephanie	Mcgee
12	SUSAN	MUELLER	Susan	Mueller
13	QUOC	NGUYEN	Quoc	Nguyen
14	MARILYN	OVIATT	Marilyn	Oviatt
15	CUONG	PHAM	Cuong	Pham
16	ALANA	PRIN	Alana	Prin
17	SARAH	PURDIN-GOLDING	Sarah	Purdin-Golding
18	JOSEPH	RANIERO	Joseph	Raniero

PROPER does not adjust to the "**Mc**" issue

(Continued on next page)...

Now what self-respecting Irishman (or Irishwoman) would like to receive a report from your department with the name **Linda Mccone** rather than **Linda McCone**. Well there is an easy fix to the problem with a little more fancy formula work using the **IF** function:



Conclusion

There you go…. A culturally sensitive **PROPER** formula that should handle all of those with "Mc" in their last name. I just made my Irish ancestors proud.

-Patrick Cummings is Director of Research and Evaluation for Tacoma Public Schools and is a regular contributor. Contact him at pcummin@tacoma.k12.wa.us

Using Assessments within an RTI Framework

By Mike Jacobsen

Background

The most recent issue of *Educational Leadership* (2008) is entirely devoted to use and misuse of data in schools today. Educators are confronted on a daily basis with a sea of data: diagnostic, and norm-referenced standardized assessment data, reading assessment data, state and district mandated assessment data. District and school administrators, teacher leaders and classroom teachers are asked to be data literate, to be able to use multiple types of assessment and other data to inform decisions that lead to higher student achievement.

What is RTI?

Response to Intervention (RTI) is an integrated, multi-tiered approach to instruction, assessment and intervention that allows schools to identify struggling students early, and provide appropriate interventions to improve student outcomes. It should be noted that RTI is a process, a way of thinking, rather then a specific program.

RTI is most commonly conceptualized as a three-tiered instructional model. This is reflected in Washington's three-tiered K-12 Reading Model. The three-tiered model is also the approach adopted by OSPI, Office of Special Education Publication: *Using Response to Intervention (RTI) for Washington's Student* (OSPI, 2006). In a three-tiered model, Tier-I (core curriculum) is the core instructional program provided to all students. At this level all students receive high quality, instruction implemented with fidelity in the general education classroom. Instruction provided in Tier-I is both differentiated and culturally responsive and is designed to serve approximately 80% to 90% of the student body. Instruction is matched to student needs. Fidelity refers to the degree to which the core instruction is implemented as designed, intended and planned. Tier-II, (strategic or supplemental) interventions/instruction are provided to the 5% to 10% of students not being successful in the core curriculum. Strategic interventions supplement the instruction being provided in the core and are targeted at identified student needs. Typically Tier-II interventions are provided in a smaller group and are carefully monitored to determine if student response is adequate. Interventions provided at Tier-III (intensive) are individually designed for the students (approximately 5%) with the most need and monitored weekly or daily for effectiveness and student response.

Core Principles

- High-Quality, Research-Based Classroom Instruction. The curriculum and accompanying instructional approaches must have a high probability of success for the majority of students. In the area of reading, for example, the core instruction needs to include the five components of the Reading First Initiative: phonemic awareness, phonics, fluency, vocabulary and comprehension.
- Universal Screening: School staffs conduct universal screening of academics and/or behavior. Specific criteria are applied to determine which students are in need of further monitoring or intervention. Many schools in Washington use DIBELS or other forms of oral reading fluency as universal screening tool.
- Progress Monitoring: This involves assessments that can be collected frequently, are sensitive to changes in student learning, and can be used to monitor a student's progress. Use of progress monitoring allows teachers to quickly identify students who are not adequately progress towards meeting standards. The frequency of progress monitoring increases as students move through the three tiered model. In the area of reading, many schools use the progress monitoring component in DIBELS. Several districts use or are developing progress monitoring procedures in math.
- Collaborative Teams: Schools develop or use existing student intervention teams to support the RTI process. The roles of existing teams usually have to change from a traditional

(Continued on next page)...

referral to special education to a problem solving process that assists with assessment and intervention design at all three tiers.

- Data Decision Rules: A common feature of RTI models is the use of data decision rules. They are applied when a student is not responding adequately to instruction/intervention and an instructional change is needed.

Assessment within an RTI Framework

How might we use assessment within an RTI framework? What kind of assessments are necessary to conduct universal, school-wide screenings? How are these assessments constructed, administered and scored? What kind of assessments are necessary for conducting progress monitoring? How are these assessments constructed, administered and scored? Related to the issue of progress monitoring, what does diagnostic assessment look like under an RTI process? Under an RTI process, assessment activities shift from summative to formative, from end of year to frequent, and repeatable. For more information on implementation of an RTI model see: (National Center on Response to Intervention- http://www.rti4success.org)

What is Universal Screening?

The first step in RTI is to accurately identify those students at risk for learning difficulties or those who are not making satisfactory progress in the Core instructional program. The screening is administered to all students. Screening is characterized by assessments that are quick, low-cost, repeatable, and test age-appropriate critical skills (e.g. identifying letters of the alphabet, decoding words in grade appropriate passages) or behaviors (e.g. tardiness, disciplinary referrals). The essential question for a screening process is whether the student should be judged as "at risk" for the target behavior. Consider how school staff use the Snellen eye chart. Using the eye chart, school staff quickly screen all students for potential vision problems. If a problem is identified by this low-cost, quick method, the student is referred for further in-depth assessment. A universal screening in reading, math or writing functions in the same fashion.

Typically in an RTI model, universal screening is conducted three times per year. In the White River School District, we conduct our benchmark universal screening in reading during the second week of September, the second week of January and the third week in May. Conducting the universal screening three times during the year allows benchmark standards to be established for fall, winter and spring and allows for the typical growth rate to be established. Using the typical growth rate allows for the comparison of growth rates of students who may be identified as at risk.

Relationship of CBM to Universal Screening and Progress Monitoring

Curriculum-based measurement or CBM (CBM,--) is a method of monitoring student educational progress through direct assessment of academic skills. CBM offers a number of advantages over other assessment methods.

- Quick to administer. For example to obtain a CBM in reading fluency, the instructor asks the student to read aloud for 60 seconds.
- Can be given often. CBM probes can be given repeatedly in a short span of time.
- Sensitive to short-term gain in academic skills, CBM has been found to be sensitive to short-term student gains.
- Low-cost. Because the assessment materials are free or low cost, and the administration time is short, the cost per student is far less than other methods.

CBM procedures have extensive research to support their use and have been developed for monitoring basic skills in reading, mathematics, spelling and writing. The initial goal for the development of CBM was to give educators simple, accurate and efficient indicators of student achievement. For more information see: National Center on Progress Monitoring http://www.studentprogress.org
Research Institute on Progress Monitoring
http://progressmonitoring.org
http://dibels.uoregon.edu/

CBM in reading When using CBM to measure oral

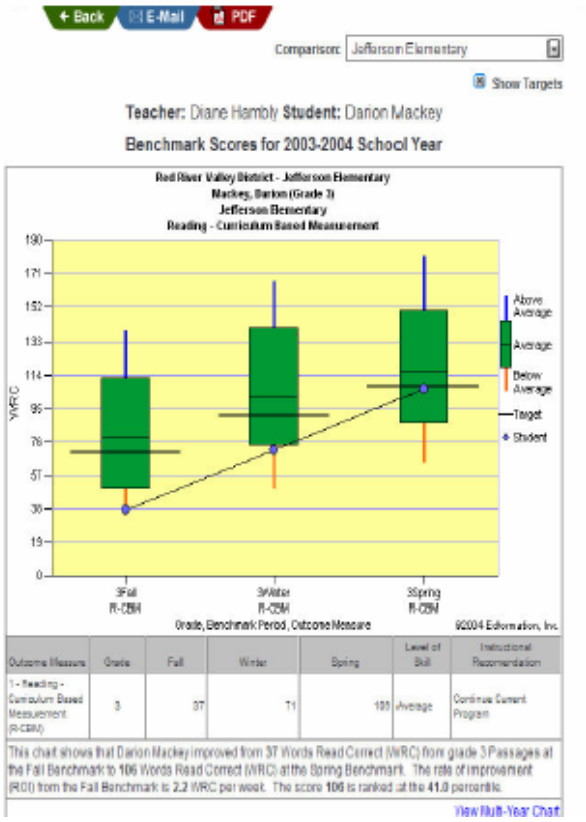
(Continued on next page)...

reading fluency (ORF) the examiner asks the student to read aloud for 1 minute from a grade appropriate passage. Using standardized directions, the score is the total of words read correctly for the 1 minute timing. Typically for use in universal screening, the student reads three passages and the median rate of words read correctly and incorrectly are the two raw scores obtained for each student. Many districts in the state and the country have adopted the DIBELS oral reading fluency directions and scoring criteria. All of the various CBM materials discussed in this section are free from the DIBELS website- http://dibels.uoregon.edu
To conduct the universal screening, staff will download the benchmark probes and follow the scoring directions. Benchmark assessments are offered by DIBELS in reading for grades kindergarten through six. The assessments are: Initial Sound Fluency, Letter Naming Fluency, Phonemic Segmentation Fluency, Nonsense Word Fluency, Oral Reading Fluency, Retell Fluency and Word Use Fluency. There is also a fee based part of DIBELS that is a data collection and management component.

AIMSWEB is another web-based CBM resource. AIMSWEB offers reading, spelling, writing and math benchmark and progress monitoring assessments and extends through grade 8. AIMSWEB is fee-based. Schools and district can purchase different options depending upon which areas they want assessed. The assessments in reading are Initial Sound Fluency, Letter Naming Fluency,

Letter Sound Fluency, Phonemic Segmentation Fluency, Nonsense Word Fluency, Oral Reading Fluency, and a Maze reading comprehension assessment. Figure 1 is a sample classroom report from AIMSWEB. The format is a box and whisker chart style. This format allows an individual student to be compared to the class. On the figure, median is shown with the line in the center of the box. The range of average reading scores (between the 25th and 75th percentile) is outlined by the box. If this was printed in color this box would be blue. Scores above average (75th to 90th) are shown by the thick vertical in at the top of the box. If this was printed from AIMSWEB, the color would be blue. Below average scores (10th to 25th) are shown by the vertical line at the bottom of the box and would be red in AIMSWEB. Readers will note that the particular student in question on this graph, started the fall with an oral reading fluency rate at the 10th percentile rank, moved to the 25th percentile rank by winter and by the spring benchmark period had moved into the benchmark range, very near the 50th percentile.

Figure 1
Sample classroom report AIMSWEB



CBM in math When using CBM to measure math, the examiner may administer probes individually or in groups. Typically, the CBMs in math consist of either single-skill worksheets that contain a series of similar problems or multiple-skill worksheets containing a mix of problems that require different math operations. The student is asked to complete as many items as possible during a 2 minute period. In the official directions, CBM in math is scored for each individual correct digit. In AIMSWEB, basic math skills are assessed via computation skills. Students are given grade level math probes with mixed math computation functions and then asked to complete as many as possible. The White River School District (WRSD) has developed an in- district math screener

(Continued on next page)...

that was implemented prior to using AIMSWEB for reading. The WRSD math screening has 20 items per grade level. 12 of the items are computational problems and 8 of the items are application problems. The math screener is also scored differently than the typical math CBM. It is scored on the basis of the number of correct problems and not the number of correct digits. It is used however in a similar fashion to the typical CBM universal screener in that it is given to all students 3 times per year, grades 1to 9.

CBM in spelling When using the spelling CBM, the examiner reads aloud words that students are asked to spell correctly within a time limit, typically 2 minutes. The spelling words are scored for correct letter sequences. Correct letter sequences are pairs of letters in a word that are placed in the proper sequence. AIMSWEB scores spelling by the number of correctly spelled words during a 2 minute timing. The use of CBM in spelling either as a universal screener or progress monitoring is much less frequent among state school districts than other CBM assessments. It's more frequent use is in the progress monitoring of special education services and in determining goals and objectives.

CBM in writing When using the writing CBM, the examiner presents the student with a story starter. The student is usually give 1 minute to think of a response and then is given 3 minutes to write the story. There are several ways the written story could be scored including the total number of words written, and number of correct word sequences. AIMSWEB scores writing either by counting the total words written or by counting correct word sequences. As is the case with spelling, the use of CBM in writing either as a universal screener or progress monitoring is much less frequent among school districts in the state, then the other CBM assessments. It's more frequent use is in the progress monitoring of special education services and in determining goals and objectives.

What is Progress Monitoring?

Progress monitoring refers to a process of ongoing data collection on academic skills of interest. The use of progress monitoring has two major purposes: (1) to determine whether students are benefiting from the instructional program, and (2) to build more effective programs for students who are not responding to instruction. Progress monitoring typically uses the same type of assessments used in universal screening. However the frequency of administration increases significantly. Instead of being administered three times per year, progress monitoring is conducted on a more frequent basis, including monthly or weekly assessment. Research has demonstrated that when teachers use progress monitoring for instructional decision-making purposes, students achieve more, teacher decision-making improves, and students tend to be more aware of their performance (Fuchs & Fuchs, 1997).

Frequency of progress monitoring increases as the student moves up the three-tiered intervention model. Although progress monitoring may be used in Tier-1, it is more likely to be used in Tier 2 and Tier 3. Progress monitoring essentially provides the indication of "response" in an RTI model. The focus of progress monitoring becomes the class, a small group, and/or an individual student. According to the Oregon Response to Intervention manual (2007), progress monitoring involves the following steps:

1. Establish a benchmark for performance and plot it on a chart (e.g.," read orally at grade level 40 words per minute by June"). It must be plotted at the projected end of the instructional period, such as the end of the school year.
2. Establish the student's current level of performance (e.g., "reads 20 words per minute").
3. Draw an aim line from the student's current level to the performance benchmark. This picture represents the slope of progress required to meet the benchmark.
4. Monitor the student's progress frequently (every Monday). Plot the data.
5. Analyze the data on a regular basis, applying decision rules (e.g., "the intervention will be changed after 6 data points that fall below the aimline").
6. Draw a trend line to validate that the student's progress is adequate to meet the goal over time.

The WRSD uses AIMSWEB to provide progress

(Continued on next page)...

monitoring probes and universal screening measures. In AIMSWEB, the progress monitoring assessments are generated electronically. For most academic areas there are enough progress monitoring assessments to be given weekly. Progress monitoring assessments can also be generated free of charge through use of the Intervention Central www.interventioncentral.org or DIBELS websites. A critical component of progress monitoring is data display. A basic understanding of Excel is helpful in generating a graphic display of the progress monitoring data. Intervention Central offers free graphing capacity through a program called: "Chart Dog-2.0" Staff will need to have progress monitoring results available and will enter the data into a program that will result in a daily or weekly graph of results. Typically a graph of progress monitoring data will list the weeks of instruction on the horizontal axis and assessment on the vertical axis e.g. correctly read words per minute. Analysis of results is significantly enhanced when data are graphed. Trend lines (graphic indication of a student's overall slope of progress) are necessary to determine whether progress is sufficient to meet the goal.

Figure 2

Reading trend line

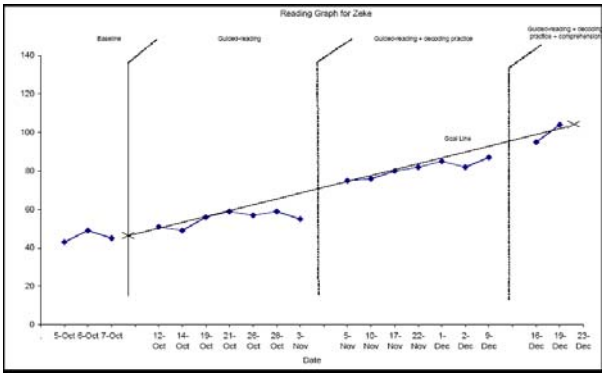


Figure 2 is one example of such a graph. The vertical axis notes words read correctly during a 1 minute timing. The horizontal access notes the weeks of instruction. An added feature of this graph is the ability to identify the different interventions that occurred over the weeks of instruction. Readers will note the interventions identified in the top portion of the graph and the addition of an aimline. The aimline is determined by the RTI team.

The application of data decision rules is an essential component of progress monitoring. Mellard and Johnson (2008) note that in order for an RTI system to be effective, several decision rules must be established. These include the following:

a. Establishing baselines. (How many data points will be needed?)
b. Establishing high but reasonable goals. (How much progress can we expect?)
c. Deciding when to make an instructional change. (Guidelines on this vary from 3 to 6 data points below the aimline.)
d. Deciding when to consider movement to another tier of intervention (either up or down).

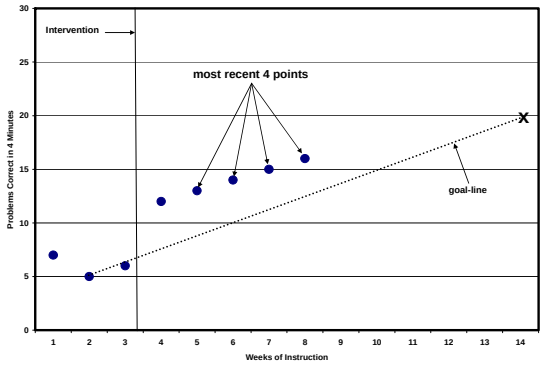
The standard used in White River is three data points for establishing a baseline. Goals are set based upon information about the student and the intervention. Benchmark goals for a given grade for fall, winter and spring provide information about how ambitious a given goal is for a particular student. An instructional change is indicated when 4 data points fall below the given aimline for a particular student. Movement across the three tiers of intervention is based upon how well the student is responding to a particular intervention. If the student is demonstrating an adequate response to intervention, it is very likely that the intervention will continue. If lack of response is observed, given the decision rules noted above, and the intervention has been implemented with fidelity, then the student may be considered for a more intensive intervention.

Figure shows several of the components outlined by Mellard & Johnson (2008). In this case the vertical axis notes the number of correct digits in 4 minutes. The horizontal axis identifies the weeks of instruction. The graph reflects three data points that established the baseline and identifies the target goal for the student at the conclusion of the 14 week intervention. The graph also identifies when the intervention began and the four most recent data points. In this particular scenario the RTI team should consider a change in the target as the student has had four data points that are above the goal-line or aimline.

Figure 3

(Continued on next page)...

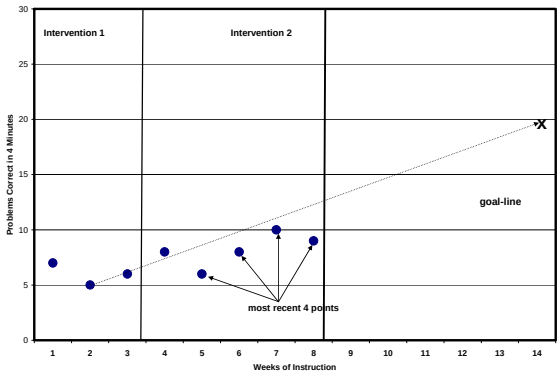
Math goal line charting



Consider Figure 4. In this particular scenario a baseline was established and an intervention was begun during the third week. However, unlike the scenario above, this particular student, with the exception of the fourth week, did not demonstrate an adequate rate of progress. All four of the most recent data points were below the aimline. Using the decision rules established in WRSD, an instructional change was indicated. The team decided to use a second intervention at week eight.

Figure 4

Math goal line charting



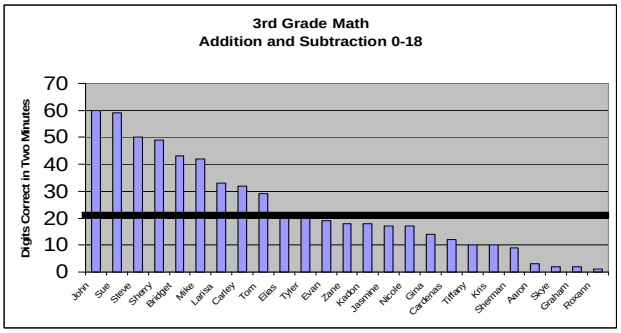
There is considerable research underway at multiple locations across the country on how many data points are needed to make an empirical decision about lack of response to intervention (National Research Center on Learning Disabilities, 2005). There is some data suggesting that increasing the number of data points to 7 or 8 does add increased validity to the decision making process. However it should be noted that adding additional data points could potentially delay implementation of a needed instructional change. The 4 data point decision rule is used extensively across the country and is supported by the research.

Use of Universal Screening and Progress Monitoring in Program Evaluation

Although much of the initial development and application of CBM used in universal screening and progress monitoring was to measure the growth of individual students, it has a strong program evaluation application. Use of these measures is increasingly seen in many school districts. It is often the case that a building has implemented an RTI model to better serve individual students and has then used the data to make system changes. Consider Figure 5. In this classroom, 14 students (56%) of the class scored below the benchmark target of 20 correct digits in two minutes. Recall that in an RTI model, 80% of students should be successful in the core instructional program in Tier-1. The classroom teacher would need to review the instructional program. In this example, it would be difficult without investigation of the instructional program, to determine which students were in fact at risk.

Figure 5

Classroom math data



If this data noted above reflected the reality in large numbers of classroom, schools and districts typically would not have the needed resources to intervene individually with every student below the target. So the focus of the intervention becomes the educational environment. Is there a core program, based upon research that is being implemented with fidelity? Is the pattern noted in the above classroom being observed in other classrooms in this particular building? Are teachers providing opportunities for differentiation, according to individual student needs? Is there frequent monitoring of student

(Continued on next page)…

performance? Is reading instruction being provided in each of the five elements of reading, i.e. phonological awareness, phonics, fluency, vocabulary and comprehension? Is sufficient instructional time being allocated for reading and math? Is the curriculum aligned with relevant content standards? These are some of the critical qC core.

Another use of universal screening and progress monitoring data is in the systematic evaluation of tiered, RTI models. Reductions in students receiving intervention/instruction in Tiers 2 and 3 is goal of an RTI approach. Use of the universal screening and progress monitoring data could be used to provide data about the percentage of students at any given tier. Often schools and buildings discover in the initial implementation of an RTI model that results to not match the RTI ideal. What is frequently revealed through the use of universal screening is substantial numbers of students at Tiers 2 and 3. Instead of data identifying 80% of students in Tier-1, implementation of universal screening might identify 60% of students at Tier-1, 10% of students in Tier-2 and 30% of students in Tier-3. These results suggest a thorough investigation of the instruction being provided in the core at Tier-1. Washington's RTI Initiative calls for data collection and analysis of the percentage of students in each Tier.

Implementation of an RTI model requires frequent and ongoing use of assessment data. This data is used to make informed decisions about the instructional needs of individual students and effectiveness of instruction and intervention being provided at all three tiers.

Annotated Resources/References

AIMSWEB may be accessed at http://www.aimsweb.com Some resources available for no charge. Most of the resources are fee based. Extensive resources for universal screening, progress monitoring data collection and display across all CBM content areas

Bender, W. N. & Shores, C. (2007). *Response to Intervention: A Practical Guide for Every Teacher*. Thousand Oaks, CA: Corwin Press.

Big Ideas in Beginning Reading: Institute for the Development of Educational Achievement may be accessed at http://reading.uoregon.edu Extensive reviews and data on research based reading programs and instruction are provided.

Buffum, A., Mattos, M. & Weber, C. (2009). *Pyramid Response to Intervention: RTI, Professional Learning Communities, and How to Respond When Kids Don't Learn*. Bloomington, IN: Solution Tree.

Curriculum Based Measures explanations may be accessed at http://cehd.umn.edu/pubs/researchworks/CBM.html .

DIBELS may be accessed at http://dibels.uoregon.edu. Resources for RTI and Interventions include the universal screening and progress monitoring materials available at no charge. A data collection and management system is available for a fee.

Educational Leadership. (2008). Vol. 66, No. 4 The entire issue from the Association for Supervision and Curriculum Development focuses on these matters.

Fuchs, L. S., Fuchs, D. (1997). Use of curriculum-based measurement in identifying students with disabilities. *Focus on Exceptional Children,* 30(3), 1-16.

Hirsch, S. & Bolz, E. & Wilson, T. (2009). *A classroom teacher's guide to RTI Assessment*, a presentation at the OSPI January 2009 Conference, Seattle.

Intervention Central may be accessed at http://www.interventioncentral.org. Resources for RTI and interventions are provided.

Mellard, D., F., & Johnson, E. (2008). *RTI: A Practitioner's Guide to Implementing Response to Intervention.* Thousand Oaks, CA: Corwin Press.

National Center on Response to Intervention may be accessed at http://www.rti4success.org . They provide comprehensive resources for RTI and interventions.

(Continued on next page)...

National Center on Student Progress Monitoring may be accessed at http://www.studentprogress.org The provide resources for RTI and interventions.

Northwest Regional Educational Laboratory, 2008, RTI*: Tiered Instruction Goes Mainstream,* Vol. 14, No .1

Oregon Department of Education. (2007).*Oregon Response to Intervention*, *Identification of students with Learning Disabilities under the IDEA 2004* may be accessed at http://www.ode.state.or.us/initiatives/idea/rti.aspx.

Research Institute on Progress Monitoring may be accessed at http://progressmonitoring.org . They provide resources for RTI and interventions.
-(2006). *Using response to intervention (RTI) for Washington's Students.* Olympia, WA: Office of Superintendent of Public Instruction.

-Mike Jacobsen is Assessment/Curriculum Director & RTI Coordinator in the White River School District. Contact him at mjacobse@whiteriver.wednet.edu.

Program Evaluation Studies by UW Graduates

Three district test directors presented program evaluation findings at the Winter 2008 State Assessment Conference in Seattle. The trio, all UW doctoral graduates working in schools, had presented the prior year as a Festschrift for their colleague and mentor, UW Prof. Catherine Taylor, to honor her work. This year Feng-Yi Hung, Clover Park, presented results of a Reading First program evaluation. Jack Monpas-Humber, Shoreline, explored the validity of Spokane's district developed assessments and convener Peter Hendrickson, Everett, explored a technique for measuring the fidelity of implementation (FOI) of new curricula. Discussant was recently UW doctoral graduate, Pete Bylsma, an independent policy and program evaluation consultant.

A parallel symposium with WSU program evaluation graduate students led by Prof. Michael Trevisan (a UW doctoral graduate and former WERA Board member) was accepted but a travel freeze nixed the presentations.

Papers from the three UW presenters are featured in this issue.

-Editor

Aligning District and State Assessments to Measure Growth in Achievement

By Jack B. Monpas-Huber, Ph.D.

Why, as depicted by the WASL results in Table 1, do students seem to be less proficient in mathematics after third grade?

Table 1
*WASL Mathematics Proficiency Rates, Grades 3-7, 2006-2008**

	Spokane			State		
	2006	2007	2008	2006	2007	2008
Grade 3	66.7	74.3	75.2	64.2	69.6	68.6
Grade 4	62.4	62.8	60.7	58.9	58.1	53.6
Grade 5	57.9	63.7	69.2	55.8	59.5	61.2
Grade 6	54.0	57.9	55.9	45.9	49.6	49.1
Grade 7	44.4	53.9	52.4	48.5	54.6	50.5

*Values are percents of students meeting or exceeding the state standard.

When Spokane Public Schools examined these results in the spring of 2007 and asked this question, it immediately looked for an *instructional* explanation: In the aggregate, were Spokane's district curriculum and instructional practices less aligned with state standards beyond 3rd grade? Were students not getting the learning experiences they needed to achieve the state standards for math proficiency?

For answers, Spokane turned to data from its own district assessments. Like many districts, Spokane had implemented a system of district interim benchmark assessments designed to measure students' mathematics achievement several times within the year prior to WASL. While many districts use commercially available assessments such as NWEA-MAP, Spokane had committed to developing its own district assessments in order to build assessment capacity and deep understanding of state standards. Consistent with the district's mission of alignment between state and district expectations, these assessments were WASL-like *by design*. They were developed using the WASL test and item specifications. They were, however, smaller than the WASL in order to be administered, scored, and the results reported back quickly. In spite of this difference in size, there was good reason to assume that the district assessments were measuring the same domain of knowledge, skills, and abilities as the WASL and reporting consistent information.

Were the data from these district assessments showing the same apparent decline in math achievement after third grade? If so, why? Were teachers teaching the district curriculum with *fidelity*? And to the extent that they were, was the curriculum *rigorous* enough to adequately move students to the state standards?

Many other districts probably asked similar questions of their district assessments about the effectiveness of their instructional programs. Such questions are fair when the primary purpose of state assessments is to provide feedback to districts and schools about the effectiveness of their instructional systems, and when districts invest in district assessment programs in order to have "multiple measures" of student achievement and rely less on one test given once per year.

Validity of District Assessments

Data from district assessments, combined with data from the state assessment, offer promise for good progress monitoring and program evaluation. However, how districts interpret and use data from multiple measures also raises important issues of validity and technical quality that districts would be wise to consider. Any time educators use any kind of assessment data to make decisions about students, such data need to be valid and reliable so that people can trust that the results are stable and measuring what they are intended to measure (Messick, 1989).

For Spokane, such validity questions about the district assessments were at least as important as the instructional implications of the results. Were

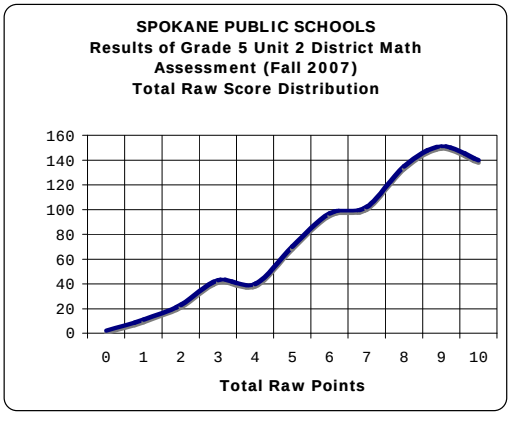
(Continued on next page)...

the district assessments reporting consistent information as WASL, and would they be predictive of WASL performance? The fact that the district assessments and the WASL were developed from the same test maps and item specifications provided strong evidence of *content validity*. But to what extent were they really measuring the same thing? What counts as evidence? At stake was no less than whether the district assessments were reporting true achievement of state standards.

The 5th Grade Mathematics Study

To try to answer some of these questions, Spokane conducted a longitudinal study. The district selected a sample of approximately 1,000 students to study over the course of their fifth grade year and ultimately into their sixth and seventh grade years. The sample was a stratified random sample of 10 students from each classroom to ensure that the sample represented the broader population of fifth grade students. Besides WASL data from grades 4 and 5, the district collected all data from the district math assessments on these fifth grade students. Ultimately, the study amounted to 10 waves of math data on these students. Item-level data were collected in order to examine the qualities of the items as well as the tests.

Figure 1
Results of Spokane District Math Assessment (Fall 2007)



The descriptive results of the study were interesting. The results shown in Figure 1 are typical of the results from the district assessments. Data from the district assessments tended to show negatively skewed distributions in which most students earned near-perfect scores. This pattern of results was good news for several reasons. It suggested that most students were learning the standards that were measured by the assessment. It also suggested that the teachers were teaching the district curriculum and that the tests were sensitive to instruction. However, these results raised other questions and implications. Did a maximum score on the district assessment really mean a student had mastered the standards measured by the assessment and would bring the same ability to the WASL? Was the curriculum rigorous enough? Such questions were cause for serious discussion among Spokane curriculum and assessment personnel.

Gathering Evidence of Construct Validity

How stable were the results of the district assessments? Were they measuring the same constructs as the WASL? At the same time that curriculum specialists were analyzing the results of the district assessments, assessment personnel were analyzing the reliability and validity of the assessments in order to answer questions like these. Two primary issues emerged which other districts that have developed in-house assessments may wish to consider.

One issue is reliability. In classical test theory, reliability is often expressed as a single statistic—Cronbach's coefficient alpha—which is based on test length and redundancy of items. Spokane's district assessments—probably like most locally developed district assessments—were necessarily short in order to administer, score, and report results more quickly. However, shorter tests are in general less reliable. Reliability becomes an issue when the total test score matters for some purpose such as correlation or prediction. A test that does not correlate very strongly with itself will not correlate very strongly with anything else. Lower reliability also means the total test score reflects other factors besides true math ability and will fluctuate if the test is administered multiple times.

A second issue is dimensionality. Classical test theory assumes that a test measures only one dimension, such as math computation. Arguably this

(Continued on next page)...

is less of an issue in reading where students can be asked to perform similar kinds of skills but with more sophistication to comprehend more challenging texts. It is more of an issue in mathematics when students are asked to perform very distinct kinds of operations or use different kinds of content such as algebra and geometry. It was an issue for Spokane's assessments insofar as each district math assessment measured a somewhat different domain of state standards than the previous, and each assessment was designed to measure several state standards. Possibly this was not so uncommon among locally developed district assessments. Dimensionality becomes an issue when one wants to use the total test score to make an inference about student mastery of a particular domain but the total test score reflects several dimensions rather than one and possibly dimensions that were not intended. Factor analysis can be a useful tool for assessing what dimensions the items seem to be measuring *based on the data* rather than the test developer's *a priori* assertions of what the items are measuring. It was not uncommon for factor analyses of Spokane's district assessments to show items clustered on dimensions other than those they were intended to measure based on the test map.

Districts that choose to develop their own in-house districts assessments and are serious about technical quality might therefore want to consider exploring these issues. They may want to consider writing focused tests designed to measure one primary dimension of learning rather than multiple, and using multiple items to measure the same performance expectation rather than one item to measure several different performance expectations.

Linking District Assessments to State Assessment

Another issue to consider with locally developed district assessments is how the assessments are *scaled*. Large-scale assessments such as the WASL are provided a scale in order to report consistent information about difficulty and student ability each year despite inevitable differences in the difficulty of different test forms and student abilities each year. Equal interval scales also facilitate arithmetic operations and statistical analyses.

Like most districts that invest in district assessments to measure state standards, Spokane wanted to make inferences from its assessments about students' performance on the WASL. One challenge that stood in the way was different scales. The WASL used the familiar equal interval scale with 400 as the proficiency standard, while the district assessments used the total raw score. One way to overcome this challenge was to put the district assessments on the same scale as the WASL so that they shared 400 as the same level of proficiency. After taking a district assessment, students could receive a scale score with 400 as WASL proficiency. The question then became how to *link* or *equate* scores from district assessments to WASL so that they share the same scale. This question prompted a review of the literature on scaling and equating (Dorans, Pommerich, & Holland, 2007) which revealed a variety of different approaches.

One promising approach to scaling uses the Rasch model, which makes use of information from the items as well examinee variation in total test scores. Bond and Fox (2001) provide an accessible introduction to the Rasch model and its applications to a variety of measurement issues. This piece includes a chapter on equating scores of the same students (a "common person" design) from two different tests designed to measure the same construct. Districts that choose to develop in-house assessments to emulate the state assessment might want to consider exploring this literature to scale district assessments in their own right or to link them to the state assessment.

Broader Issues for Districts

Districts that invest in district assessment systems for purposes of program evaluation face an important choice between two approaches. One is to use outside instruments such as MAP. The other is to develop in-house district assessments that emulate the state assessment.

(Continued on next page)...

The advantage of locally developed assessments is the development of assessment capacity as teachers become "students of the standards" (Eaker, 2008). Local district personnel develop deep understanding of state standards and how to collect credible evidence of student achievement of those standards. If the assessments are developed according to the same test map and item specifications as the state assessment, they have evidence of *content validity*. However, one challenge of this approach is the time and expense of training and freeing teachers to do this work. Locally developed assessments may also not be subject to the same rigorous validity studies for evidence of technical quality. It may also be difficult to link scores from the district assessment to the state assessment in a clear way.

Commercially available district tests, such as MAP have a different set of issues. A huge advantage of these kinds of assessments is that they likely enjoy the benefit of a very large item bank, which is beneficial in several ways. The items likely enjoy very high quality and have known difficulty and discrimination statistics based on piloting. This makes possible computer adaptive testing which helps provide more reliable measurement of student abilities. Such tests also enjoy the benefit a continuous scale of growth for measuring student achievement across grade levels. However, it may be more difficult for these kinds of instruments to claim *content validity* when they are not developed according to the same test map and item specifications as the state assessment.

References

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences*. Second Edition. Mahwah, NJ: Lawrence Erlbaum.

Dorans, N. J., Pommerich, M., & Holland, P.W. (Eds). (2007). *Linking and aligning scores and scales*. New York, NY: Springer.

Eaker, R. (2008). Statement made in conference session on the role of common assessments in professional learning communities. Seattle, WA.

Messick, S. (1989). Validity. In *Educational Measurement*, R. Linn (Ed.). Washington, DC: American Council on Education.

-Jack B. Monpas-Huber, is Director of Assessment and Student Information Shoreline Public Schools and held similar postions in Spokane and Northshore. He serves as Secretary to the National Association of Test Directors.

Curriculum Renewal: Fidelity of Implementation

By Peter Hendrickson, Ph.D.

Much effort, expense and time are spent in the process of textbook selection, sometimes cast within the framework of program renewal, more often as the purchase of new texts. Underlying is a Theory of Action with the belief that different (improved) student outcomes are possible if there are new materials. Typically a year or so is spent in the selection (or program recasting) phase with the following year devoted to implementation of the new text or program.

While professional development is often the center piece of the new text/program, the entire implementation is sometimes not planned or supported as thoroughly as is the selection process. Program Evaluation is a process used to determine if the new program is being implemented with fidelity (FOI) and if the impacts observed are the impacts desired.

This paper focuses on the FOI component asking the question, "To what extent is the enacted program consistent with the intended program model?" (Century et al, 2007). Suggested steps include:

- Develop a Theory of Action or Logic Model to display the relationship of the context, activities, measures and outcomes. Given the presenting problem, how do the planned actions lead to the outcomes? How will you measure those outcomes?
- Identify critical components in four areas:
    - Structural/procedural--says what to do in the classroom, the most basic steps of the procedures of instruction and the physical organization of the program to make that clear to the teacher. **Example**: The Read 180 model calls for a room divided into three areas for each of the instructional components.
    - Structural/educative--tells what the teacher needs to know to use the program as intended. What is the basic level of content and pedagogy needed? **Example:** Elementary teachers may be shy of solid conceptual knowledge. They may need instruction to themselves instruct ratio and proportion.
    - Instructional/pedagogical--tells what instructional strategies teachers will use. **Example:** The Read 180 model calls for brief large group instruction, small group instruction/modeling, independent or guided reading, software time and whole group wrap up. Do instructors know how to use guided reading?
    - Instructional/student engagement--outlines expectations for student engagement. **Example**: Students read 18 books at or above their independent level each year.
- Determine not only the degree to which the program is implemented as intended but describe the principal variations in implementation. It is widely recognized that programs are rarely implemented precisely as designed at the school or classroom level. Acknowledging and describing the variability gives the opportunity to measure the impacts of alternate models.
- Measure the implementation using both quantitative and qualitative methods.

How will the evaluator know if the novel curriculum is being implemented with fidelity?

Direct Observation

While direct observation of curriculum implementation is the norm, observation tools may not be available. Teacher evaluation protocols governed by negotiated contracts may be an obstacle to direct observation as program evaluation activities may be suspected of being teacher evaluations in disguise. If teacher coaches or instructional

(Continued on next page)...

facilitators are asked to conduct the observations, they risk eroding their position of trust. If contracted (external) observers from outside the district are employed, teachers may feel the observations are not being conducted in service of aiding the teacher, rather they are advancing administrative needs.

Several measures may be taken to enhance the utility of the direct observations:

- Develop an observation protocol in collaboration with teachers.
- Provide the protocol to teachers prior to the observations.
- Be clear about the intended use of the data from the observations. Do not deviate from the stated use.
- Train the observers in use of the protocol.
- Provide timely feedback to the teacher following the observation.
- Give the teacher the opportunity to provide information about the implementation beyond the protocol bounds.
- Demonstrate how the observation data is being used, ad hoc, to improve program.

Following is a partial model for observation of curriculum implementation which may be used as a template for direct observation. A fully realized model is contained in Appendix "E" from the University of Chicago math and science critical elements work (CEMSE, 2008). This abbreviated example is intended for use with a new handwriting curriculum.

Table 1

Framework for curriculum renewal fidelity of implementation.

Component	Not Present	Partial	Adequate	Substantial Compliance
What to do				
Procedures	• New curriculum shelved; old curriculum yet in use • Only taught sporadically • Does not use scoring rubric	• New curriculum enhanced by or enhances original curriculum • Taught less than 15 minutes daily • Occasional use of scoring rubric	• New curriculum the heart of many lessons; used for planning • Taught at least 15 minutes 3X/week • Scores most samples with rubric	• New curriculum in use; old curriculum absent • Taught daily 15 to 20 minutes • Integrates scoring rubric into all lessons
Physical Org	• Materials not at hand • No student samples	• Many students have materials • Work samples, no exemplars	• Most students have materials • Work samples, exemplars	• All students have materials • Refreshed samples, exemplars
Needs to know				
Content	• No knowledge of new program	• Knows only initial content	• Familiar with all components	• Mentors others on content
Pedagogy	• Misunderstands methods	• Spotty background information	• Familiar with strategies for each component	• As needed, displays new strategies

(Continued on next page)...

Instructional Strategies				
Prior learning	• Ignores skills in hand	• Stretch zone not found regularly	• Most students work at right level	• Builds on prior learning for each child
Direct instruction	• Workbooks replace D.I.	• D.I. either too long or short	• D.I. regular but pacing irregular	• Lessons are explicit, clear
Guided practice	• Does not monitor work	• Sporadic work monitoring	• Each lesson provides guided practice	• All students practice at level
Assessment	• Instructive feedback absent	• Assessments not used to inform instruction	• Assessments becoming student centered	• Teaches self, peer assessment
Engagement				
Instructional	• Essential work not completed	• Portions of lessons completed	• Essential work is completed	• Students apply learning beyond lesson
Student	• Students off task	• Students on task parts of lesson	• Students start w/o prompting, mostly on task	• Students self-assess

Ideally, the entire population of teachers implementing the new curriculum would be observed. Practically, a sample could be drawn for observation so that inferences might be made to the population. Whichever method is used, the observers must be adequately trained in the use of the protocol so that findings across schools and classrooms are reliable. If resources permit, some observations would be either repeated later by a second observer to check for inter-rater reliability or two observers would independently employ the protocol during a single observation.

Where a spirit of collegial professionalism exists (or is under construction) peer-to-peer observation can be both powerful for building internal expectations for continuous improvement and economical for gathering FOI data. A single substitute hired for a day could provide released time for a teacher or teachers to observe several colleagues, the Teacher Expectations of Student Achievement (TESA) model (Kerman, 1979).

Several other tools are available to collect FOI data beyond a direct observation protocol. These indirect measures include:

- Surveys which ask implementers to rate the extent to which critical and other program elements are present. Web survey software makes possible fast construction, easy administration (most teachers have email accounts) and quick analysis as the software aggregates and displays results at no additional cost in time for data entry and reporting (Zoomerang, --).
- Focus groups include representative interviewees, times when most can gather, field testing the interview protocol, transcribing and validating the data, and conducting content analysis (Bamberger et. al. 2006, pp.63, 285).
- Interviews with implementers or supervisors provide a two-way vehicle to gauge the presence of program elements and learning about unanticipated circumstances of the intervention (Hendrickson, 2008.)
- Software monitoring is possible if software use is a component of instruction. At a minimum, the evaluator will know if the

(Continued on next page)...

student was present if time on program is logged.
- Student Information System data may tell if the student was present during instruction through attendance applications.

Curiously, a new textbook/materials adoption encumbers a significant district cost, while expectations for improved performance are modest. While internet book resellers advertise new algebra texts for under $5, 10 years ago high school math books topped $60 each for a new adoption. When professional development costs are included, the totals rise with the teacher per diem and presenter stipends. Districts generally do not expect rosy impact data in the first year as there is a widespread belief (Fullan, 2001) that an implementation dip, lower test scores, are to be expected as teachers learn how to instruct the new lessons from unfamiliar materials.

However, when discretionary or categorical funds are used to augment the standard curriculum with a targeted intervention, expectations are for an immediate impact. In some cases, year two carry-on program funding is dependent on year one results. Program managers who have not monitored FOI have greater exposure to program termination if impacts are neutral or negative and they can not reliably print to FOI issues as possible/probable explainers. Appendices B, C, and D reflect conversations with Everett Public Schools curriculum specialists faced with ongoing curriculum renewal and materials adoptions issues.

Curriculum adopters/renewers are part of the accountability web. If they do not plan to evaluate the fidelity of implementation, they have compromised the system's ability to understand results. Few are the programs which provide the tools for a fidelity check, but when they do, prudent program managers will use them to advantage (Placement, Assessment and Reporting Guide, 2006; Borman, et al., 2007).

Annotated References

Bamberger, M., J. Rugh, & L. Mabry. (2006). *Real World Evaluation*. Thousand Oaks, CA: Sage Publications. Three veteran program evaluators reveal tricks of the trade in running evaluations with limited resources, time or commitment.

Borman, G., Slavin, R.E., Cheung, A., Chamberlain, A., Madden, N.A., & Chambers, B. (2007). Final reading outcomes of the national randomized field trial of Success for All. *American Educational Research Journal, 44* (3), 701-731. When implementing Success for All, considerable emphasis is placed on fidelity. External evaluators, extensive training for building principals, building-wide commitment and exhaustive instructional rubrics set the program apart from most other literacy models.

CEMSE. (2007). *Fidelity of Implementation Questionnaires for Mathematics Curriculum in Grades 1-5*. , Chicago, IL: University of Chicago.

CEMSE. (2008). *Critical Components: Definitions and Explanations*. Retrieved November 21, 2008 from the Worldwide Web at http://cemse.uchicago.edu/files/CCDefinitions summary 2008_11_110.pdf
This fully fleshed math and science FOI checklist provides a useful base document for creating local checklists.

Century, J., Freeman, C., Rudnick, M., & Leslie, D. (2007). *A conceptual framework for fidelity of implementation of instructional materials.* Paper presented at the Annual Meeting of the American Educational Research Association, New York, NY. March 28th 2008.

Fullan, M. (2001). *Leading in a Culture of Change*. San Francisco, CA: Jossey-Bass.

Hendrickson, P. (2008). Unpublished principal interview protocols for several program evaluations. Everett, WA: Everett Public Schools.

- (2006). *Read 180 Placement, Assessment and Reporting Guide*. NYC, NY: Scholastic. The PARG is a companion tool to the SFA observation protocol

Kerman, Sam. (1979). Teacher Expectations and Student Achievement. *Phi Delta Kappan*, v.60, n.10, p. 716-18, June 1979.

(Continued on next page)...

National Sciences Resource Center. (1977). *Science for All Children*. Washington, D.C.: National Academy Press.

Zoomerang is available free via www.zoomerang.com on the Worldwide Web to implement surveys for limited use. A full education license is $149 per year. Survey Monkey www.surveymonkey.com offers similar services at $200 per year.

-Peter Hendrickson, is an assessment, research and program evaluation specialist with Everett Public Schools. Appendices are available upon request phendrickson@everettsd.org.

Does Reading First Work?

By Feng-Yi Hung, Ph.D.

Special thanks to the support and assistance from the Clover Park School District Compensatory Program. Ann Cuoio and her staff provided valuable information. Without their help, we wouldn't have the context and history of the Reading First program.

Introduction

As assessment and program evaluation coordinators, we know a single study of one school district is never able to capture all the information that can be gained about a particular program or initiative. Instead, it takes a systematic approach by conducting multiple studies over time to provide a thorough understanding of how well a program works and how effective the impact on student achievement the program has.

What is Reading First?

Reading First is a federal initiative to strengthen the instruction of reading to primary grade students. Since 2003, Reading First has provided a substantial amount of federal funding to states and districts for K-3 reading programs, with the goal of having children read at grade level by the end of third grade. The Reading First program includes using a research-based core reading program, hiring a reading coach, providing at least 90 minutes of reading instruction per day, regularly assessing students' reading skills and providing reading intervention to struggling students. Reading First schools, in general, have high rates of student poverty and low levels of reading achievement.

Reading First Evaluation

OSPI conducted Reading First evaluation annually by using the following qualitative and quantitative data.

1. Student assessment – K-3 scores on DIBELS.
2. Spring Surveys – paper surveys of all teachers, coaches, principals, district coordinators
3. In-person interviews – principals, coaches, 2 teachers from each school
4. Classroom observations – during site visits, targeted observations of three reading lessons at every school selected for a site visit
5. Interview with state project staff members

Clover Park School District's Results

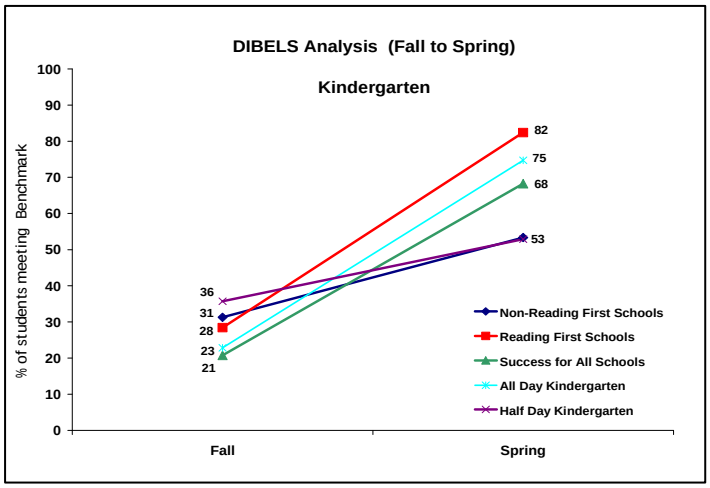
DIBELS Results

The Dynamic Indicators of Early Literacy Skills (DIBELS) was the primary measure of student outcomes in Reading First schools. This assessment includes a set of standardized, individually administered measures of early literacy development. Kindergarten and first grade include several assessments and phonemic awareness and phonics. Starting mid-year of first grade, oral reading fluency was used to measure students' reading achievement. Students obtaining adequate scores on these assessments are said to be "at benchmark," while the students scoring at the lowest level fall into the "intensive group."

The first four graphs, Figures 1 to 4, show the percent of matched students scoring at benchmark from the fall to the spring.

Figure 1

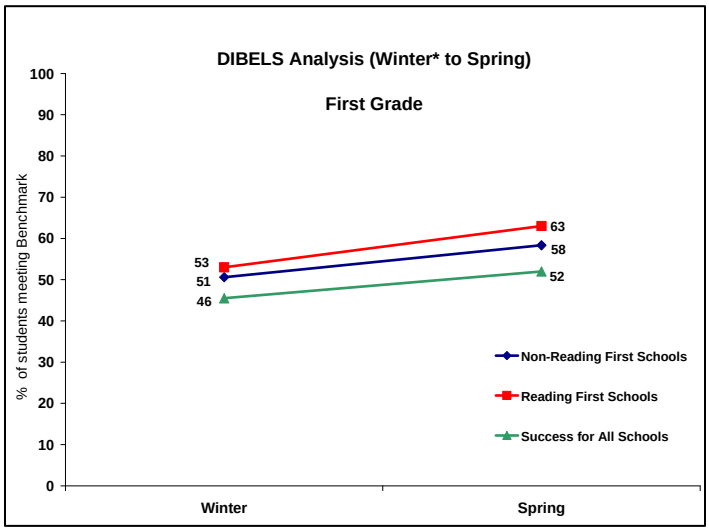
DIBELS results for kindergarten.



(Continued on next page)…

Figure 2

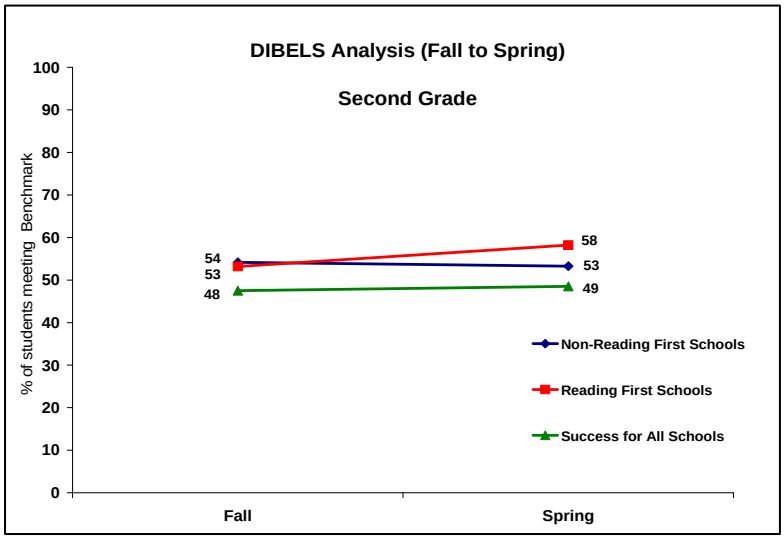
First grade DIBELS scores by program based on Oral Reading Fluency (ORF). ORF was not administered in the Fall testing session.



Non-RF schools are comparable to SFA in terms of growth. SFA is implemented in initially lower performing schools.

Figure 3

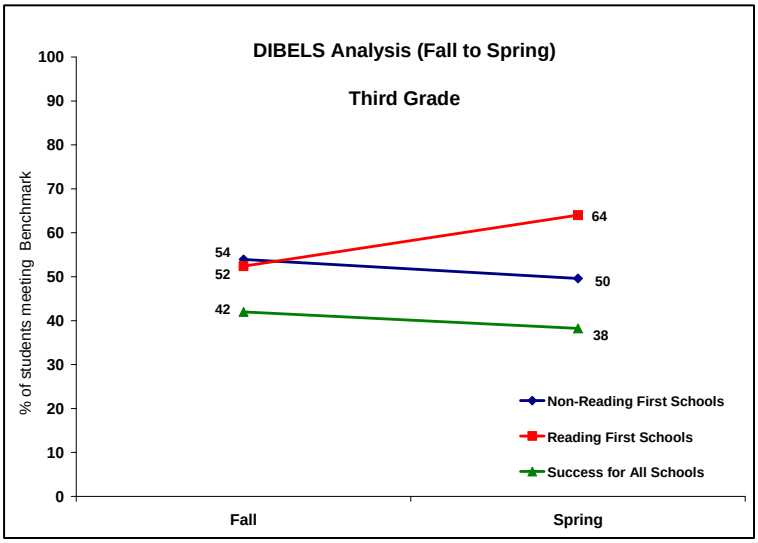
Second grade DIBELS results.



SFA is not making the progress as much as RF. Is it because SFA focuses on something more than reading fluency?

(Continued on next page)…

Figure 4
Third grade DIBELS results

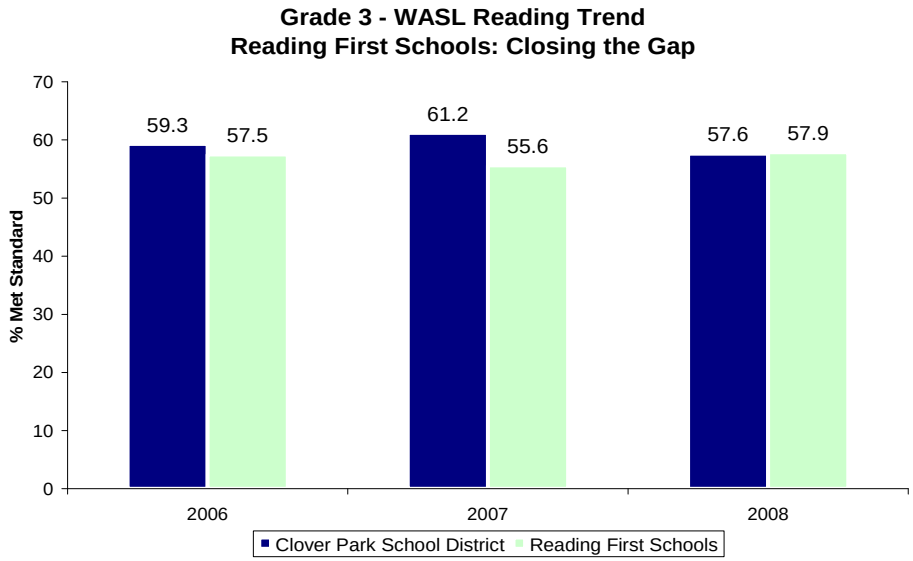


Clearly, RF schools performed much better than non-RF and SFA on oral reading fluency, as measured by DIBELS.

WASL Results

Clover Park School District has been an active participant in Reading First since 2003. In 2006, Reading First schools scored slightly lower then the district in grade 3 WASL reading; in 2008, Reading First Schools scored higher than the district. Similar trends were found in upper grades. In grades 4 and 5, Reading First schools made more gains from the beginning years of Reading First Program (2004 and 2006) to 2008. The gap between Reading First schools and the district is closing because the rest of the district is not doing well. This is not a good way to close the achievement gap.

Figure 5
Grade 3 WASL – percentage of students meeting standard for the district and Reading First schools.



(Continued on next page)…

Figure 6
Grade 4 WASL – percentage of students meeting standard for the district and Reading First schools.

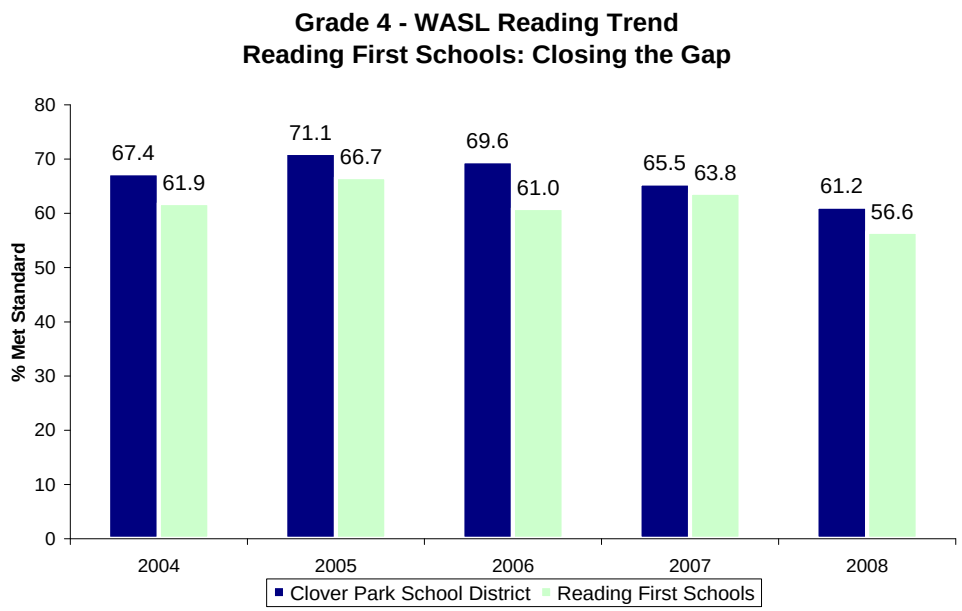
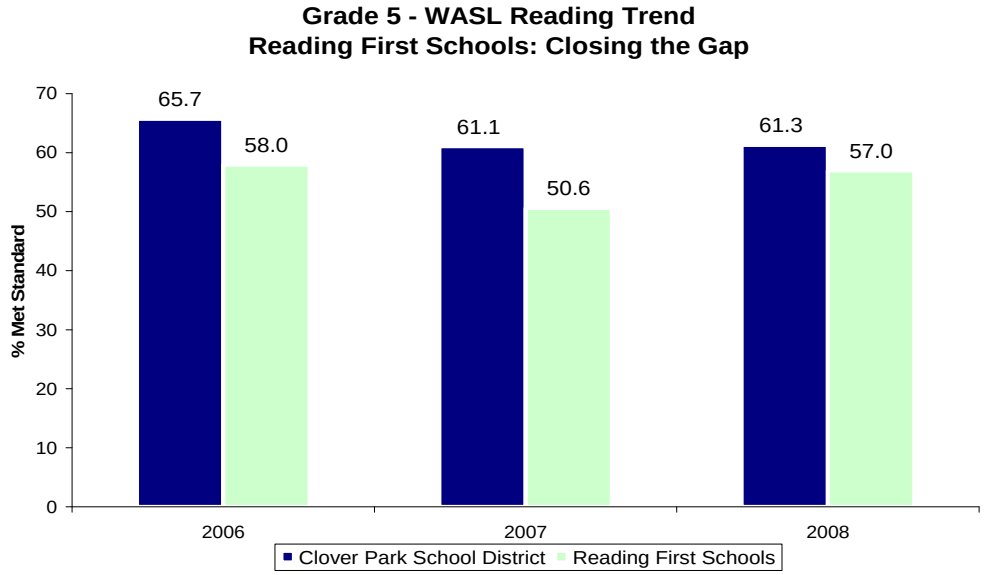


Figure 7
Grade 4 WASL – percentage of students meeting standard for the district and Reading First schools.



There hasn't been much change for RF schools. However, the rest of the district dropped by 4.4% from 2006 to 2008.

(Continued on next page)…

WASL and DIBELS Results

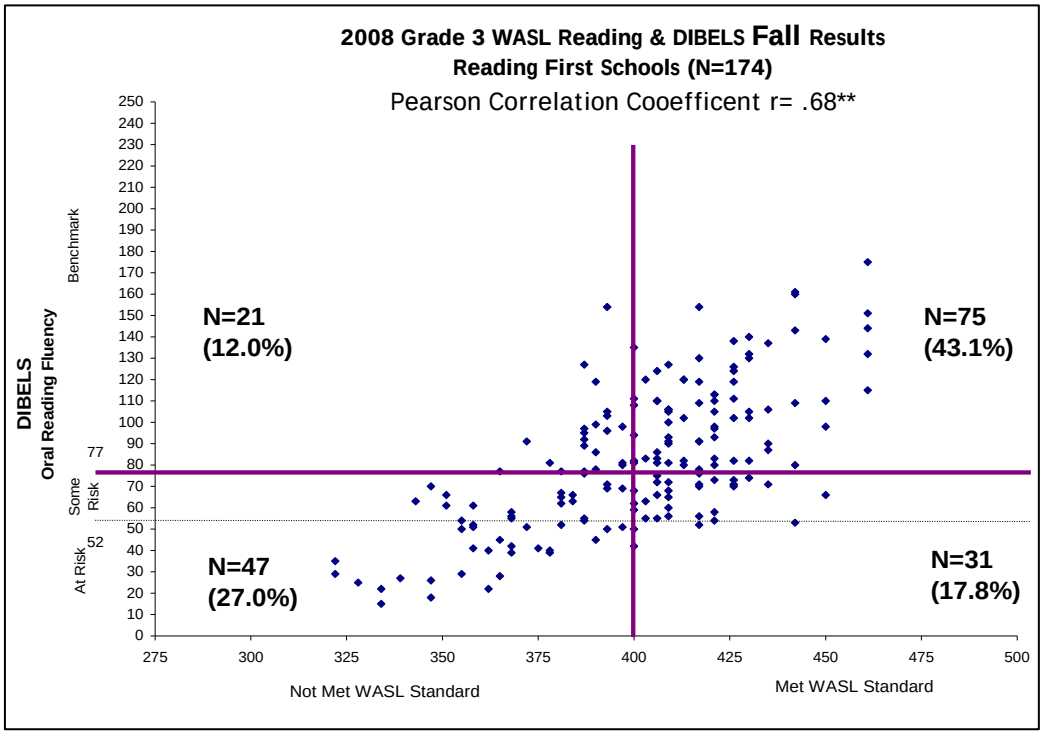
The relationship between WASL and DIBELS is interesting. Pearson correlation coefficient between Grade 3 WASL reading and DIBELS (Fall and Spring) is slightly lower than 0.70. This correlation is statistically significant at the 0.01 level (2-tailed). I assumed the correlation coefficient between the fall DIBELS results and WASL would be much lower than the spring DIBELS and WASL due to the intervention and timing of two assessments. However, this assumption was proven invalid.

Figures 8 and 9 show the relationship between WASL and DIBELS results. We found –

1. Reading First schools did a nice job of improving students' Oral Reading Fluency (ORF) by moving them out of "At Risk "and "Some Risk" to "Benchmark" for students meeting standard on the WASL (approximate 60%).
2. About one quarter of students were identified as scoring below the DIBELS benchmark and did not meet standard in the WASL.
3. About fifteen percent of DIBELS benchmark students did not meet standard in the reading WASL. These students are likely to have strong fluency in their reading performance, but they are struggling with understanding the reading passages. In other words, they are good at word-calling, but their comprehension was below the state standards. They are likely to be ELL students or minority students.

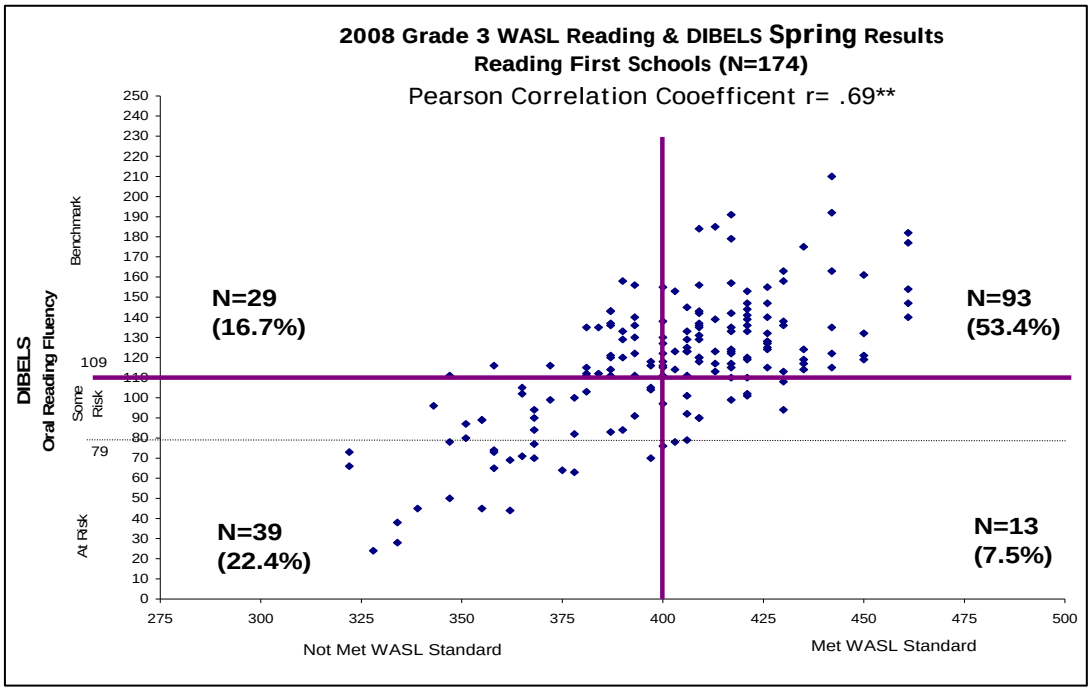
The results highlight the discrepancy of these two assessments and the gap that occurs if reading intervention is based on DIBELS scores only. We need to communicate with elementary principals and teachers about the assessment purposes for each assessment.

Figure 8
Grade 3 WASL and DIBELS Fall Results for Reading First Schools (N=174).



(Continued on next page)…

Figure 9
Grade 3 WASL Reading and DIBELS Spring Results for Reading First Schools (N=174).



Discussion

DIBELS data shows that students in Reading First schools from kindergarten to third grade outperformed other groups, such as Success For All, full-day kindergarten, half-day kindergarten and non-Reading First schools. The progress of Reading First schools, as measured by DIBELS assessment, is substantial and impressive.

However, when we use state assessments like WASL to measure the program impact, we did not find as much positive program impact as we expected. In general, Reading First schools remained the same while the rest of the district scores went down (it could be worse when both Reading First and the rest of the district had a decrease)- Reading First schools are closing the "achievement gap" slightly faster than the rest of the district, as measured by the WASL (by 1% to 4%).

The moderate correlation between DIBELS and WASL reading is interesting, but not surprising. They are two different assessments with two different goals in the same content area. In other words, we are measuring two different reading constructs. DIBELS focuses on early basic reading skills and WASL measures the state reading standards in comprehension, communication and higher-level thinking. Specifically, Grade 3 DIBELS only measures oral reading fluency (ORF) – how fast and accurate students are able to read within a very short length of time. This finding leads to questions about the usage of DIBELS results in intervention placements, identifying struggling students, and using DIBELS results to predict WASL success.

Other Questions/Future Studies

This is meant as an initial look into Reading First School performance in the Clover Park School District. I am thinking of expanding this research to all Reading First schools in the state of Washington. Since Reading First implementation occurs in different districts and each district has its unique curriculum, professional development and instruction cultures, there may be considerable differences in Reading First program impact. I would also like to do some kind of Hierarchical Linear Models (HLM) where I can examine the unique relationship of nested factors.

(Continued on next page)…

This HLM method will make it possible to separate the variance into components explaining the effects of different levels of analysis, such as the effects of school, district, years of Reading First program, quality of 90-minute reading block, ELL, SPED and socio-economic status. The interactions among these factors on Reading First program impact, as measured by DIBELS as well as WASL, will be interesting.

Currently Reading First schools, in general, follow the state and federal implementation guidelines by hiring a reading coach, providing at least 90 minutes of reading instruction per day, regularly assessing students, providing reading intervention and using a research-based core reading program. In other words, resources are provided and program structures are in place. Program fidelity seems to be there. Now the questions are – Is the quality of reading instruction satisfactory? Are reading coaches able to work with staff and have positive influence on teachers' practices? What effects in terms of student achievement are we expecting? After gathering more student achievement data (more than one district) and apply advanced regression analysis (beyond descriptive statistics), we will be better informed about the impact and effectiveness of Reading First schools.

References:

Kachigan, S. K. (1991). *Multivariate statistical analysis: A conceptual introduction.* New York: Radius Press.

Loehlin, J.C. (1998). *Latent variable models: An introduction to factor, path, and structural analysis.* 3rd Edition. London: Lawrence Erlbaum Associates.

Popham, J. (2009). Unraveling reliability. ASCD: *Educational Leadership* (pp 77-78).

U.S. Department of Education. *Frequently Asked Questions Related to the Decrease in Fiscal Year 2008 Funding for the Reading First Program (June 2008).* Retrieved on December 17, 2008 from http://www.ed.gov/programs/readingfirst/faq.html.

University of Oregon. *DIBELS data system.* Retrieved on December 15, 2008 from https://dibels.uoregon.edu/resources.php.

-Feng-Yi Hung is Director of Assessment and Program Evaluation in Clover Park School District. She was formerly a data analyst at OSPI.

Linking a Comprehensive Professional Development Literacy Program to Student Achievement

By Nancy Katims, Ph.D., N. Lynn Caulkins, Lara Drew, and Maggie Conners

Over the past several years, the Edmonds School District has utilized a model of professional development with multiple components including:

- demonstration classrooms with observation opportunities for teachers
- coaching by peers and by expert coaches
- collaborative learning teams and study groups facilitated by teacher leaders
- district sponsored workshops
- comprehensive summer institutes
- other professional development opportunities.

This combination of professional development strategies has focused on improving literacy practices of teachers in grades K-6. An important resource by Diane Sweeney (2003) has provided a foundational source of learning for the project participants. Principals and teacher leaders studied key concepts related to leading effective professional development. The group has focused on learning strategies for examining effective literacy practices as well as developing facilitation skills and using protocols to look at student work. District literacy coaches have developed their coaching skills and strategies to support school-based learning

Despite the inherent difficulties of isolating and measuring the effects of professional development on student achievement, we conducted an evaluation study to determine whether this set of professional development activities provided through the district's Collaborative Literacy Project (CLP) has a measurable impact on student performance in literacy.

Because CLP encompasses many different types of professional development activities, the researchers used a non-traditional approach to evaluating the program. Basically, they assigned "CLP PD" (professional development) points to all district elementary teachers based on the various CLP activities in which the teachers had participated over the 05-06 and 06-07 school years. The points were weighted according to the proportional value of different activities in line with the research-based activities considered most relevant to improving literacy development. For example, receiving one-to-one in-class coaching was weighted four times compared to attendance at a literacy institute.

The approximately 500 elementary teachers were then placed on a "CLP PD" point continuum. About 40 of these teachers were identified as having high enough points to have a solid foundation of CLP professional development. A comparison group of 40 teachers was chosen from teachers on the "low" end of PD points to match the student demographics of the "high" PD point classrooms as closely as possible.

Despite best efforts to match the student groups, compared to students in the Low CLP PD group, a higher percentage of the students in the High CLP PD group were from low income homes, Hispanic, and limited English proficient. Therefore, based on decades of research, students in the High CLP PD group would be *predicted* on average to perform less well on standardized measures than those in the Low CLP PD group.

We also examined the demographics of the teachers in the two groups. Compared to the High CLP PD teachers, the Low CLP PD teachers on average were more experienced in their overall years of teaching and had earned more overall graduate credits and "clock hours" in professional development. In other words, even though this group did not participate in the CLP activities, they participated in a great deal of other professional development.

Table 1 shows the student achievement measures for each of the two groups. Basically, the students of the High CLP PD teachers performed better on nine out of 10 reading measures than did the students of the Low CLP PD teachers. An interesting correlate is that the High CLP PD students did not perform better in math, showing that these students and teachers were not simply "more competent" than the Low CLP PD group, but

(Continued on next page)...

rather demonstrated higher achievement specifically in reading.

Table 1

Achievement Outcomes for Students of High and Low CLP Professional Development Teachers

Grade	*Pre-Reading and Reading Measures*	*High CLP PD*	*Low CLP PD*
K	*Spring 07 DIBELS: % of students in low risk or established category*	98%	65%
1	*Spring 07 DIBELS: % of students in low risk or established category*	77%	53%
	Fall 07 Grade 2 Reading Assessment: *% of students who met or exceeded target in accuracy/fluency* *% of students who met or exceeded target in retelling*	 51% 68%	 57% 65%
2	*Fall 07 Grade 3 District Reading Comprehension Assessment:* *% of students who met or exceeded target*	 43%	 42%
3	*Spring 07 Reading WASL: % of students who met/exceeded standard*	66%	60%
4	*Spring 07 Reading WASL: % of students who met/exceeded standard*	74%	73%
5	*Spring 07 Reading WASL: % of students who met/exceeded standard*	77%	63%
6	*Spring 07 Reading WASL: % of students who met/exceeded standard* *Mid-year 06-07 District Grade 6 Reading Comprehension Assessment:* *% of students who met or exceeded target*	80% 73%	63% 52%

Grade	*Mathematics Measures*	*High CLP PD*	*Low CLP PD*
2	*Spring 07 Grade 2 District Math Assessment:* *% of students who met or exceeded target*	 52%	 68%
3	*Spring 07 Math WASL: % of students who met/exceeded standard*	66%	64%
4	*Spring 07 Math WASL: % of students who met/exceeded standard*	53%	58%
5	*Spring 07 Math WASL: % of students who met/exceeded standard* *Mid-year 06-07 District Grade 5 Math Assessment:* *% of students who met or exceeded target*	60% 31%	50% 34%
6	*Spring 07 Math WASL: % of students who met/exceeded standard*	55%	44%

(Continued on next page)...

Grade	*Writing and Science Measures*	*High CLP PD*	*Low CLP PD*
4	*Spring 07 Writing WASL: % of students who met/exceeded standard*	57%	59%
5	*Spring 07 Science WASL: % of students who met/exceeded standard*	37%	20%

We then tried to replicate the study, looking at student achievement in the 07-08 school year. We formed two new groups of teachers, one with high points and a matched group with low points, based on their CLP PD points combined across the 05-06, 06-07, and 07-08 school years. We applied the same evaluation design as in the first study. While the results were not as strong or consistent in the second study as in the first study, the findings leaned in the same direction. Table 2 provides a summary of the results in both studies.

Table 2
Summary of Student Achievement Outcomes in the Two Studies

	Study 1	*Study 2*
***Pre-Reading and Reading** (main focus of CLP)*	The High CLP PD group outperformed the Low CLP PD group on 9 out of 10 pre-reading and reading measures.	The High CLP PD group outperformed the Low CLP PD group on 5 out of 10 pre-reading and reading measures by at least 6 percentage points, and was within 3 percentage points of the low CLP group on 4 of the remaining 5 comparisons.
***Math** (not a CLP focus)*	The High CLP PD group out-performed the other group on 3 of 6 measures, and the Low CLP PD group outperformed on 3 of 6 measures.	The High CLP PD group out-performed the other group on only 1 of 6 math measures.
***Writing** (not CLP focus until 07-08)*	The Low CLP PD group outperformed the other group by a small margin.	The Low CLP PD group outperformed the other group.
***Science** (measure that relies heavily on reading)*	The High CLP PD group outperformed the other group by a large margin.	The High CLP PD group outperformed the other group.

In evaluating the design used in this study, it is important to note the following:

- ✓ "High CLP" teachers were identified only by the number of activities in which they participated, not by actual observation of their classroom practices. The inference is that they use effective literacy practices as a result of the CLP PD activities, but we of course do not know if this is true for all the "High CLP" teachers.
- ✓ The activities tallied as CLP PD points included only those provided through district coaching staff. Some schools provided CLP training through their school staff, and could have been included in the "Low CLP" group despite their involvement through school-sponsored professional development.

(Continued on next page)…

For next steps, we plan to narrow the focus of the professional development to the activities supported most strongly in the research literature, particularly one-on-one coaching and demonstration classrooms (Barth, 2006; Neufeld & Roper, 2003). Additionally, we would like to focus on the intermediate grades, an age when reading comprehension becomes so important to school success, and on schools with high ELL populations.

For a copy of the complete report, email Nancy Katims at katimsn@edmonds.wednet.edu.

References

Barth, R. (2006). Improving relationships in schools. *Educational Leadership, (March) 9-12.*

Neufeld, B. & Roper, D. (2003). *Coaching: A strategy for developing instructional capacity.* Providence, RI: Annenberg Institute for School Reform.

Sweeney, D. (2003). Learning along the way: Professional development by and for Teachers. Maine. Stenhouse. 2003.

-Nancy Katims is Director of Research and Evaluation for the Edmonds School District and plays a leadership role in the WERA test director network.
-N. Lynn Caulkins is an Assessment and Data Specialist.
-Maggie Connors is a District K-6 Literacy Coach.
-Lisa Drew is Director of Elementary Education

Washington Educational Research Association
PO Box 64489
University Place, WA 98464

We're on the Web!

Visit us at:

www.wera-web.org

The Standard Deviation Newsletter

Editor
Peter Hendrickson, Ph.D.
Everett Public Schools
4730 Colby Avenue
Everett WA 98203
425-385-4057
phendrickson@everettsd.org

Photo Editor
Don Schmitz
Mukilteo School District

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

Humor Co-Editors
Michael Power,
Tacoma School District
Bob Silverman
Puyallup School District

Technical Co-Editors
Patrick Cummings,
Tacoma School District
Jack Monpas-Huber, Ph.D.
Shoreline School District

Editorial Assistant
Jeanne Willard
Everett Public Schools

Layout Designer
Michelle Sekulich
The Boeing Company

Executive Secretary
Leonard Winchell
WERA
Po Box 64489
University Place, WA 98464
253-564-4816
lenwwa@aol.com

The Standard Deviation is published spring, winter, and fall as an online newsletter and journal. Submissions are welcomed from WERA members and others. Kindly submit articles for consideration using APA format.
Copy deadline for Spring is May 8, 2009.