The WERA Educational Journal
Volume 2 Number 2 July 2010

i

FIG. 5.

The sixth-grade boy's drawing (Fig. 4), being rated as half way between *f* and *i*, has 2½ degrees of merit, that is, is 2½ times as far from zero as drawing *c* is. If a certain other drawing, say that of Fig. 5, is rated as one-fifth of the way from *i* to *m* it has 3.2 degrees of merit and is 1.28 times as 'good' as the drawing of Fig. 4 in the same sense[1] that a lift of 320

A Scale for the Merit of Drawings by Pupils 8 to 15 Years Old

FIG. 6, *a*.

This was intended to be a picture of a man.

o merit.

INSIDE THIS ISSUE

Editor's Corner

While state education agencies take a deep breath after submitting (or not) Race to the Top applications, we in the research and program evaluation community are reminded our role is to step back from the fray and speak truth to power, as best we understand it. This issue brings a close look at report card validity from Shoreline's **Jack Monpas-Huber** and a thoughtful approach to school enrollment forecasting by former Seattle demographer Les **Kendrick**. OSPI math director **Shannon Edwards** casts light on mathematics benchmark assessments from the Summit project and consultant **Pete Bylsma** provides an update on the new State Board of Education Accountability Index.

Ethics Editor and data analyst **Andrea Meld** introduces a forum on data and research ethics with a review of the recently released *Forum Guide to Data Ethics* including the Integrity, Data Quality and Security Canons.

Your editor reflects on the role of research in textbook selection in the light of local legal action in the "math wars." He also reviews a web-based approach to research and program evaluation in the face of reduced evaluation resources across districts. Book Review Editor **Phil Dommes** offers a trio of reviews by OSPI's **Andrea Meld**: *Applied Longitudinal Analysis*, Kent's **Bob Isenberg**: *Information Dashboard Design*, and Tacoma Housing Authority's **Michael Power**: *The Trouble with Diversity: How We Learned to Love Identity and Ignore Inequality* provide choices for summer reading.

Researchers and their colleagues are provided a checklist from the AERA graduate student forum to self-evaluate the strength of their work or provide feedback to other researchers. Finally, "novice foodie" and Journal editorial assistant **Heather Bandeen** of OSPI reviews Tacoma's Rosewood Cafe, a personal favorite. Heather applied her considerable editorial skills to much of the copy in the current issue to better match APA style and journal conventions. We're aiming towards a suggested citation at the foot of each article and abstracts in future issues.

The cover art comes from Photo Editor Don Schmitz who has been exploring old psychological measurement texts and graphics in Google Books. *Thorndike Graphic link*.

Northwest education colleagues are invited to submit assessment, research and program evaluation articles, letters to the editor, and opinion pieces for consideration in further issues. Copy deadlines are October 1 for the late fall 2010 issue and March 4, 2011 for the late spring issue. All submissions should follow APA format and include the full name, affiliation and contact information for the writer. Letters may be edited for space.

--Peter Hendrickson, Ph.D.

AERA Conference Notes: Program Evaluator in a Can? Not Quite.

Dollars are tight everywhere across the educational landscape, even with targeted federal grants at unprecedented levels. And an accountability culture remains on the rise.

Against a backdrop of sinking funding and rising expectations for achievement, district resources (read staffing) for research and evaluation are finding new lows across the land. When I walked into the annual meeting of the national Directors of Research and Evaluation group at the American Educational Research Association (AERA) annual conference two months ago in Denver, I was surprised to know each of the 30 or so in attendance--not a fresh face in the lot. And four more of us had retired last June. The news was the same from Miami to Chicago and Los Angeles to St. Paul--researchers and program evaluators are disappearing from district offices.

Talk around the circle turned to getting the evaluation jobs done with contracted services. Several East Coast colleagues mentioned Hanover Research in D.C. and I took in a session a few days later by a West Coast outfit, Empirical Education from the Bay Area. Empirical has evaluated a math interim assessment project, one of my current evaluations areas as an OSPI evaluation contractor.

I was struck by their promotion of an online program evaluation tool, MeasureResults ®, and wondered how an "evaluator/researcher in a can" would work for districts which had shed (or never had) program evaluators on staff. A few weeks later I scheduled a webinar to see how a district might "conduct research at a low cost, in an effective manner by outsourcing the analytic and reporting function." (Empirical Education, 2010). The software was developed with federal funding to provide an inexpensive method for schools to conduct their own research with an online web tool. Working through the software demonstration was like wading through TurboTax ®, but without the deductions for donated statistics texts.

The software was helpful in gathering essential data in a logical sequence, but limited in scope and nuance. My fiscal life is complicated enough that a CPA is needed despite our careful filing and spreadsheets. I suspect most program evaluations start with simple questions (Did this work?) but the details very quickly generate complexity and unforeseen issues. Further, the details of evaluations are often an after-thought.

The designers have created a rule-driven template to guide the district evaluator in selecting a design and providing data in a prescribed format followed by viewing a short menu of reports. The webinar presenter was careful to note that this new tool is not an entire evaluation, but part of a portfolio of evidence. Fidelity of implementation, for example, is not addressed. While surveys are possible, qualitative studies are beyond the capability of the tool. And some of the most important work remains asking the right questions.

It will be interesting over the years ahead to see if the evaluator corps returns or even grows. In the meantime, many districts will be looking at contract services to help with critical evaluation projects. We plan to provide a review of regional contractors in an upcoming issue.

References

Empirical Education. (2010). Retrieved May 8 from http://www.empiricaleducation.com/news2010.php

--Editor

Validity Of Elementary Standards-Based Report Card Grades

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

In his influential chapter, Samuel Messick (1989) defined validity as "an integrated evaluative judgment of the degree to which theoretical rationales and empirical evidence support the adequacy and appropriateness of inferences and actions based on test scores and *other modes of assessment*" (p. 13, emphasis mine).

In education, it is common to question the validity and reliability of scores from standardized tests, especially when test scores carry high stakes for students. But what about those *other modes of assessment*—such as report card grades? Do we hold report card grades to the same standards of technical quality as we do standardized tests?

Report card grades have been a part of American public education for a century, and for about as long, observers have questioned their validity as measures of student achievement. As early as 1913, Starch and Elliott questioned the reliability of grades in English and mathematics, pointing out variability between teachers in what skills they grade on and how they choose to grade. Now it is well-documented that teachers assign final report grades to students based upon a variety of factors other than student achievement. Additionally, teachers vary in the way they approach assessment and grading (Guskey & Bailey, 2001; Marzano, 2000; O'Connor, 2007).

Such variability comes at a cost. As students of comparable skill and achievement receive different report card grades and people will inevitably make decisions upon the basis of those grades. For example, at the elementary level, some people may look at high grades in elementary school as evidence that a student should be placed in a gifted program, or an accelerated middle school mathematics course. Others may view low grades as evidence that a student should be placed in an instructional support program. In high school, grades carry even higher stakes as these subjective low grades may cost students credits and crucial time--or even admission to a selective college or university. Such decisions may be construed as risky if based upon information that arbitrarily measures factors superfluous to student achievement.

In recent years, many districts and schools have attempted to reform grading practices by establishing "standards-based" report card systems (Guskey & Bailey, 2001; Trumbull & Farr, 2000). These reporting systems aim to overcome the inherent limitations of traditional letter grades and teacher-based report cards in several ways. One is by establishing a standard report card template used by all teachers and schools across all grade levels in a district, effectively doing away with teacher- or school-specific report cards. Often this template includes a standard rubric or rating scale—possibly modeled after performance levels on the state assessment—to report the achievement of students at the end of a grading period. This template may also list a variety of skills in each content area outlined as important by state standards. Together, these features probably bring at least some technical consistency to bear on assessment and grading practices across districts. In addition, the recent literature on grading does much to educate teachers gently about validity issues with grades and provides some methodologies to improve their assessment and grading practices. However, with some exception (Randall & Engelhard, 2009), comparatively less has been written about the psychometric qualities of report card grades—especially elementary standards-based grades—as measures of student achievement. Are the new standards-based grades reliable? Do they correlate with scores from the state assessment that they are designed to predict?

The purpose of this study is to examine the psychometric properties of elementary standards-based report card grades. Specifically, I investigate the criterion-related (predictive and concurrent) validity of a recent sample of fourth grade mathematics grades by examining their associations and correlations with scores from their primary criterion, the *Washington Assessment of Student Learning (WASL)*. I find evidence of a reasonably strong relationship between these measures and discuss the significance of this finding for educational measurement and practice.

Data and Method

The data for this study come from 625 fourth grade students in a greater metropolitan Seattle area school district. Mathematics grades from the 2008-09 school year were extracted from the student information system. Grades were posted at the conclusion of each of three trimesters. The fall trimester concluded at the end of November; the winter trimester in mid-March; and the spring trimester in mid-June. These student grade records were then matched to mathematics scores from the state assessment taken in April. In this particular district, students receive grades in specific mathematics skills and not an overall mathematics grade. For this reason, if an overall or summative subject area grade was required for statistical analysis, I calculated the modal trimester math grade.

Criterion-Related Validity of Elementary Standards-Based Report Card Grades

If the purpose of the elementary standards-based grades is to report student achievement of the skills expected by the state standards, then these grades should bear some reasonably strong empirical relationship

Continued on next page

to scores from other measures of the same skills. The extent to which grades correlate with criterion scores in the state assessment is evidence of *criterion-related* validity (Allen & Yen, 1979). There are two types of criterion-related validity evidence. When grades precede the state assessment scores in time, this is called *predictive* validity. When grades and scores are generated at approximately the same time, this is called *concurrent* validity. This section presents analyses of the association between grades and scores on the state assessment for each of the three trimesters.

Table 1 presents the association between the modal fall math grades and those same students' performance on the *WASL* that spring. For ease of interpretation, the values along the diagonal (where grades and assessment scores correspond) are bolded, and lines are used to demarcate grades and scores defined as "meeting standard" from those that are below standard.

A look at the scores above the *WASL* standard of Level 3 suggests a fairly strong association between grades and state assessment scores. Looking at grades and scores of 3 or above, nearly all students whom teachers declared as proficient on local standards (receiving mostly 3s and above on their report card) actually went on to meet or exceed state proficiency standard months later. Similarly, very few students who received mostly grades of 1 on their report card went on to meet the state standard. Most of the errors in prediction concerned the 149 students who received mostly 2s on their report card. Only about half of these students met the standard. This suggests that by December, teachers were fairly sure about students achieving at the extreme ends of the scale but less certain about students who were close to local proficiency standard.

Table 2 presents the relationship between the modal winter math grades and state assessment scores. The difference here is that these grades were posted in mid-March, closer to the state assessment in April. This closer timing would suggest the winter grades should be even more predictive of state test scores than the fall grades.

The association is fairly strong. At the high end, nearly all of the students receiving grades of 4 actually met the state standard, most of whom at Level 4. At the lower end, nearly all of the students receiving grades of 1 later scored below the state standard. Teachers are fairly accurate judges of performance at the extreme ends, while once again it is the students in middle of the distribution who present more challenge to classify. About eight percent of students whom teachers gave mostly 3s did not meet the state standard, a slightly less accurate prediction for students at this performance level over the fall grades. Of the 136 students who received mostly 2s on their report card, slightly under half actually met the state standard. This is about the same as the results for the fall grades.

Table 1
Modal Fall Trimester Mathematics Grades and WASL Performance Level

	Modal Fall Trimester Math Grade				
WASL Level	1	2	3	4	Total
1	**15**	36	7		58
	71.4%	24.2%	1.7%		9.5%
2	4	**37**	21		62
	19.0%	**24.8%**	5.1%		10.2%
3	2	59	**188**	4	253
	9.5%	39.6%	**46.1%**	13.3%	41.6%
4		17	192	**26**	235
		11.4%	47.1%	**86.7%**	38.7%
Total	21	149	408	30	608
	100.0%	100.0%	100.0%	100.0%	100.0%

Table 2
Modal Winter Trimester Mathematics Grades and WASL Performance Level

	Modal Winter Trimester Math Grade				
WASL Level	1	2	3	4	Total
1	**15**	35	10		60
	62.5%	25.7%	2.6%		9.9%
2	6	**35**	21	1	63
	25.0%	**25.7%**	5.5%	1.5%	10.3%
3	3	51	**190**	10	254
	12.5%	37.5%	**49.6%**	15.2%	41.7%
4		15	162	**55**	232
		11.0%	42.3%	**83.3%**	38.1%
Total	24	136	383	66	609
	100.0%	100.0%	100.0%	100.0%	100.0%

Continued on next page

Table 3 presents the results for the modal spring grades and state assessment performance. Unique about these grades is that they represent the final grades posted in June, after students have already taken the state assessment. As a result it makes less sense to think of them as predictors of scores on the state assessment than as concurrent information. Yet, these spring results are similar to those from the previous two trimesters as most of the error in prediction occurred among students in the middle of the distribution. Of the students graded as just below local standard (received mostly 2s on their final spring report card), approximately 42 percent had actually met standard on the state assessment.

Finally, Table 4 reports the Pearson correlations between the average math grades and scale scores on the state assessment. The correlations are moderately strong, indicating good evidence that the two measures are capturing the same construct. However, it is interesting that the fall and spring grades are slightly more predictive than the winter grades, which are closer to the time of the actual state testing.

Table 3
Modal Spring Trimester Mathematics Grades and WASL Performance Level

	Modal Spring Trimester Math Grade				
WASL Level	1	2	3	4	Total
1	**14**	38	11		63
	56.0%	29.7%	2.9%		10.2%
2	8	**37**	18	1	64
	32.0%	**28.9%**	4.7%	1.3%	10.4%
3	3	42	**199**	13	257
	12.0%	32.8%	**51.7%**	16.3%	41.6%
4		11	157	**66**	234
		8.6%	40.8%	**82.5%**	37.9%
Total	25	128	385	80	618
	100.0%	100.0%	100.0%	100.0%	100.0%

Table 4
Correlations between Mean Grades and WASL Scale Scores

Mean Trimester Grade	N	Pearson correlation (r)
Fall	625	0.643
Winter	625	0.628
Spring	625	0.707

Discussion

The purpose of this study was to examine the criterion-related validity of a sample of elementary standards-based report card grades by exploring their relationship with scores from the state assessment. The preceding analyses of the three sets of trimester grades demonstrate fairly consistent results. The report grades—at least for fourth grade mathematics, in this year—were fairly strongly predictive of performance levels on the state test, especially at the extreme ends of the two distributions. There was less correspondence in the middle of the distribution, especially among students who were achieving slightly below proficiency levels. At this level, teachers seemed to under predict actual performance on the state assessment. It could be that teachers err on the side of caution with predictions about these borderline students. This may also be a population of students who are highly capable but whose grades reflect behavioral or other non-achievement kinds of issues. Overall, these grades showed moderately strong correlations with test scores. Arguably, the size of these correlations is evidence of reliability, for if a measure is unreliable it will not correlate with anything else. This measurement research carries several implications for educational practice. The first is that districts should regularly examine the technical quality of their report card systems, in part through analysis of correlation, then share this information with schools and teachers so they can see how closely their local expectations and professional judgments about student proficiency align with state expectations. This analysis could provide grist for good discussion among teachers at the same grade level about what exactly it means to "meet standard" in a content area both within a district within a grading period as well as on the state assessment in the spring.

Districts may also support teachers in this effort by providing not only the time and forum for collegial discussion but also by developing and providing *operational definitions* of report card grades. These definitions would guide teacher grading as well as help parents understand what each grade means in terms of concrete skills. More commonly held operational definitions between teachers and across districts of what exactly students should know and be able to do in a content area at a grade level, and what counts as evidence of proficiency, would greatly benefit students, especially those who move between schools.

Continued on next page

References

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Prospect Heights, IL: Waveland Press, Inc.
Guskey, T. R., & Bailey, J. M. (2001). *Developing grading and reporting systems for student learning*. Thousand Oaks, CA: Corwin Publishers.

Marzano, R. J. (2000). *Transforming classroom grading*. Alexandria, VA: Association for Supervision and Curriculum Development.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: Macmillan.
O'Connor, K. (2007). *A repair kit for grading: 15 fixes for broken grades*. Portland, OR: Educational Testing Service.

Randall, J., & Engelhard, G., Jr. (2009). Examining teacher grades using Rasch measurement theory. *Journal of Educational Measurement, 46*(1), 1-18.

Starch, D., & Elliott, E. C. (1913). Reliability of grading work in mathematics. *The School Review*, *21*(4), 254-259.
Trumbull, E., & Farr, B. (Eds.). (2000). *Grading and reporting student progress in an age of standards*. Norwood, MA: Christopher-Gordon.

--Jack Monpas-Huber is a frequent Journal contributor and currently serves as President of the National Association of Test Directors. He is Director of Assessment and Student Information for Shoreline Public Schools. Contact him at jack.monpas.huber@shorelineschools.org.

Tweaking a Cohort Forecast

By W. Les Kendrick, Ph.D.

Many techniques have been recommended to improve school enrollment forecasts (Choi and Wang, 2006; Grip, 2004) but, after many years, cohort survival continues to be the most widely used method for forecasting K-12 enrollment by school administrators (Shaw, et.al., 1997; Deckel, 1994; Schellenberg & Stephens, 1987). The popularity of the technique is partly attributable to its simplicity. Cohort survival focuses upon the net change in enrollment as students progress from one grade to the next. For example, second grade enrollment in a given year is compared to 1st grade enrollment from the previous year to assess the amount of growth or decline that occurs over the school year. The trends at each grade level can be averaged over several years and applied to the current year enrollments to predict future enrollment.

Although cohort survival is widely used, research suggests that projections are often improved by combining different methods (Armstrong, 2001; Clemen, 1989). This could involve averaging the results of different methods or adjusting the results of a cohort model to take advantage of additional information. For the present discussion, I would like to present a typical cohort survival approach to projecting enrollment with a discussion of some issues that arise in using this method. I would then like to suggest some other methods and data that can be used to tweak a cohort forecast. For purposes of the present discussion, we will limit ourselves to a forecast of five years, which allows us to avoid the additional problem of predicting births. For a five-year forecast, we typically know how many births have happened in the previous five years that are likely to impact enrollment in the coming five years.

Table 1 shows the enrollment by grade level for a hypothetical school district between 2005 and 2009. To the right of 2009 enrollment is the cohort average for each grade, which indicates the average net change in enrollment as students progress from one grade to the next. In all cases I have used a five year average to predict future enrollments. The average is obtained at first grade, for example, by summing the 1st grade enrollments from 2005 to 2009 and dividing them by the sum of the kindergarten enrollments from 2004 to 2008. The kindergarten enrollment is then multiplied by the cohort rate (1.069) to get a first grade projection for 2010. For kindergarten, the table compares kindergarten enrollment to county births from fuve years prior to the enrollment year to create an average birth-to-k ratio. This is the district's average share of the county birth cohort. This ratio is multiplied by the birth cohorts eligible for school in the coming years to create a kindergarten projection. The numbers are rolled forward to subsequent years by taking each grade level projection and multiplying it by the appropriate rate.

Table 1
Typical cohort survival model (5 year average)

Birth Year	1999	2000	2001	2002	2003	2004			Births	2005	2006	2007	2008	2009
County Births	7596	7595	7734	7316	7580	7695		Cohort		7914	8002	8010	8044	8369
								Survival						
	Enrollment Year							Rate		Projection				
	2004	2005	2006	2007	2008	2009		5 Yr Avg		2010	2011	2012	2013	2014
K	951	876	925	938	967	955	*	0.123	K	973	984	985	989	1029
1	927	1043	932	998	1000	1004		1.069	1	1021	1040	1052	1053	1057
2	933	930	1045	937	1028	1017		1.012	2	1016	1033	1052	1064	1065
3	953	922	951	1059	964	994		1.004	3	1021	1020	1037	1056	1068
4	955	946	939	946	1026	980		0.998	4	992	1019	1018	1035	1054
5	1004	970	956	929	981	1030		1.011	5	991	1003	1031	1030	1047
6	998	1012	975	978	943	993		1.013	6	1043	1004	1016	1044	1043
7	1048	1022	1038	955	1024	924		1.011	7	1004	1055	1015	1027	1056
8	1025	1048	1031	1044	965	1031		1.006	8	930	1010	1062	1021	1034
9	1125	1022	1076	1048	1026	1004		1.013	9	1044	942	1023	1075	1034
10	1054	1090	1006	1043	1025	1019		0.979	10	983	1022	922	1001	1052
11	934	975	1059	953	985	950		0.943	11	961	927	964	869	944
12	927	934	999	1073	1023	1055		1.036	12	985	996	961	999	901
Total	12834	12790	12932	12901	12957	12956			Total	12964	13055	13138	13263	13384

*Kindergarten is the average share of the birth-to-K ratio

One question that is frequently asked is whether one should use a straight average, or a weighted average that weights recent enrollment years more heavily than others. Whether you use an average or a weighted average will depend upon whether you think recent trends reflect a shift in enrollment that is likely to endure for the next five years, or whether you believe recent trends represent random variation from the average of a typical pattern. A similar issue is at play when people ask about how many years to use in calculating an average (three, five, six years, etc.). People typically opt for fewer years when they want to emphasize the most current trends. But a recent study commissioned by Washington State, found no substantial overall differences in accuracy when comparing three year and five year cohort models for producing a five year projection for all districts in the State (OSPI, 2008).

Research also suggests that we can be fooled by the psychological effect known as "anchoring," in which we tend to assume that recently received information is predictive of the future, when in fact the future may look quite different from the present (Tversky & Kahneman, 1974). How many people, for example, were surprised by the decline in the real estate market that began in the summer of 2007?

These issues alert us to a major quandary that comes with using cohort survival. In practice the method is a pure mathematical formula and makes no assumptions about what is causing the trends. Knowing what average to use requires knowing something about

Continued on next page

the demographic trends that are affecting enrollment (e.g. births, home school or private school enrollment, housing growth, movement into and out of existing neighborhoods). And it requires knowing how trends might change in the future. Short of hiring a consultant to analyze these trends, there are some things that districts can do to provide some perspective.

For example, one might reasonably ask if future housing and population growth is likely to be greater or less than what has been seen in the last few years. Most cities and counties can provide information about the number of single family and multi-family permits for a specified time period. And in the Puget Sound and Spokane areas, the company, New Home Trends, tracks the permitting and sales of new homes by school district. If you can obtain an estimate of the number of homes that were permitted or sold in the past five years in your district area (comparable to the years used in the cohort model) and an estimate of the number of units permitted for the most recent year (likely to impact enrollment over the next few years), it is possible to compare these two numbers and use the difference to adjust the forecast. Let us assume, for example, that a district sees about 50 students per year for every 100 new homes that are built (most districts see between 40 and 60 students from new single family homes). If there are 200 new homes per year, this particular district would see about 100 students from those new homes. But if future development drops to a rate of 100 new homes per year, the district will see 50 fewer students per year. The cohort projection can thus be adjusted down by 50 students a year to account for less growth from housing.

Another way to adjust a cohort forecast is to use estimates and forecasts of the school-age population. The Office of Financial Management at the State of Washington provides this information on their population web site for every county. The steps for making this adjustment are as follows:

1) Estimate the annual percentage growth in the county school age population for the cohort history period (2005-2009) using the State data which shows the population estimates for ages 5-17. (Past school-age growth rate)

2) Use the growth management forecasts from the State and get the school-age forecast for your county for the next five years (2010-2015). Calculate the annual percentage growth that is expected. (Future School-Age Growth Rate)

3) Use the difference between these two percentages to adjust your cohort forecast up or down, whichever is appropriate. For example, if the future growth is expected to be 3-tenths of percent higher on an annual basis than past growth, you can adjust the cohort forecast at each grade by this amount.

Because this method uses county level data, it may not be as sensitive to specific housing or other trends that are present in your district. But remember that the cohort forecast already indicates the average change in enrollment that has been present in your district for the historical period. The adjustment based upon the county school age trends is an attempt to better align predicted growth with predicted future trends in the school-age population. This approach is a good alternative for districts that are not highly impacted by new housing development, or where data on housing is not readily available. It also provides an alternative to the straight cohort approach.

There are many other methods available to tweak or check a given cohort forecast, including regression models that predict enrollment using births, housing and population forecasts, and population/ratio models. These variables relate the size of your school district's enrollment to the size of some larger population. You can also rely on the personal judgments of local experts, although some argue that personal judgments are not as reliable as specific methods, nor as easily corrected when they turn out to be wrong (Grove and Meehl, 1996; Armstrong, 1983). Whatever the methods used, it is good to remember that all forecasts contain error. Taking the average of forecasts produced with different methods, tweaking a cohort forecast to produce a different result, and considering low and high alternatives during the planning process can help you avoid putting all your forecast eggs into a single cohort basket.

References

Armstrong, J. (2001). Combining forecasts. In J. Armstrong (Ed.) *Principles of forecasting: A handbook for researchers and practitioners*. Norwell, MA: Kluwer Academic Press.

Armstrong, J. (1983). Relative accuracy of judgmental and extrapolative methods in forecasting annual earnings. *Journal of Forecasting* 2, 437-447.

Choi, S. & Wang, P. (2006, August). *Using GIS and modeling to improve small area school enrollment projections*. Paper presented at the 2006 ESRI International Users Conference, San Diego, CA.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting 5*, 559-583.

Deckel, G. P. (1994). Projecting school enrollment: A solution to a methodological problem. *School Business Affairs, 60* (4). 32-34.

Grip, R. (2004). Projecting enrollment in rural schools: A study of three Vermont school districts. *Journal of Research in Rural Education, 19* (3). Retrieved March 2010 from http://www.psu.edu.

Grove, W. M. & Meehl, P. E. (1996). Comparative efficiency of informal (subjective impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law*, *2* (2), 293-323.

Continued on next page

Office of the Superintendent of Public Instruction, Olympia Washington (2008, December). *K-12 enrollment forecasting study*. Olympia, WA: Department of Printing. Retrieved (August 2009) from EnrollProjectionMethodologies

Schellenberg, S. J. & Stephens, C. E. (1987, April). *Enrollment projection: Variations on a theme.* Paper presented at the annual meeting of the American Educational Research Association, Washington D.C.

Shaw, R. C., Alspaugh, J., & Wat-Aksorn, P. (1997). Accuracy of enrollment forecasting methods. *ERS Spectrum, 15* (4), 16-19.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science 185*, 1124-1131.

--Les Kendrick was the demographer for Seattle Public Schools from 1990 to 1997. Since 1997 he has been a consultant to local school districts providing demographic analysis and enrollment projections, and assisting districts in the management of student assessment data to track student learning. He can be reached by e-mail at les@nweds.com.

Summit Mathematics Benchmark Assessments: Linking Student Learning to System Alignment

By Shannon Edwards

Background

Mathematics interim or benchmark assessment development tabled by OSPI during the budget crisis two years ago has flowered across several districts in a partnership with OSPI and an external developer.

The District and School Improvement and Accountability (DSIA) Division partners with eight school districts in Washington State and four external partners in the Summit District Improvement Initiative. Summit districts develop knowledge, skills, and capacity to lead and support consistent, sustained and dramatic increases in student achievement. As Summit districts accept the challenge of dramatically raising student achievement, Summit educators assist in the development and field testing of improvement tools and strategies that can be scaled to other districts.

The Summit Mathematics Benchmark Assessments were developed in 2009 with an external partner, Teachscape, to provide urgently needed support for districts to use state standards to align mathematics curriculum and guide instruction. During the 2009-2010 school year the assessments are undergoing field testing in the eight Summit districts. Table 1 (above right) illustrates the number of students assessed by district.

It was important that DSIA's purpose for the Summit Mathematics Benchmark Assessments be clearly articulated and communicated to district leaders. The Summit Mathematics Benchmark Assessments are interim assessments whose purpose is to provide a bridge between classroom formative assessments and the end-of-year summative assessments, and provide information to:

Table 1
Approximate numbers of students tested per grade band and district

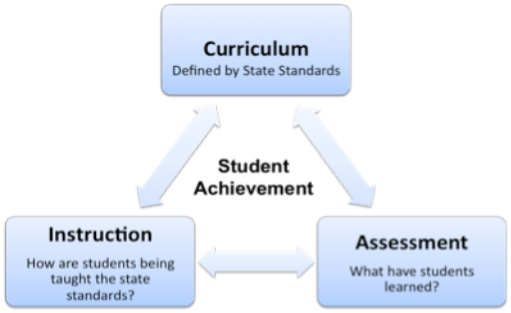
	K-2	3-5	6-8	High School
Clover Park	2125	2642	2266	1472
Mt. Vernon	1364	1218	1111	984
Mt. Adams	270	210	195	143
Othello	851	817	692	642
Renton	3201	3277	2533	2043
Sunnyside	1403	1346	1292	952
Tukwila	666	593	548	548
Wapato	195	427	649	567
Totals	**10,075**	**10,530**	**9,286**	**7,351**

- Evaluate student learning of specific state mathematics standards
- Identify patterns of student need to inform changes to the mathematics program and instruction through collaborative dialogue
- Identify specific student misconceptions or errors as they relate to the content in the specific standards being assessed

Rationale for Interim Assessments

We know when students experience an aligned system of curriculum, instruction, and assessment tied to state standards, student achievement improves (see Figure 1). In high poverty, high performing schools the curriculum is aligned and implementation is monitored (Barr & Parrett, 2007). When looking closely at the improvement strategies of these high poverty, high performing schools, effective use of interim assessments is often a key ingredient (Marshall, 2006). However, implementation matters. Tools and structures for the analysis and use of the interim data to inform instructional decision-making prove to be essential to the effectiveness of the interim assessments.
Office of Superintendent of Public Instruction, link to OSPI page

Figure 1
Relationships between curriculum, assessment and instruction



Continued on next page

With new state mathematics standards the Summit Mathematics Benchmark Assessments were developed as a tool to help districts monitor the implementation of these new standards in the context of district curriculum and pacing guides. Equal access to rigorous content is strongly related to student performance (Schmidt & Cogan, 2009). Therefore, it is important that (1) the district curriculum and pacing guides align with grade level standards, (2) instruction focuses on mastery of standards and (3) the benchmark assessments assess the standards previously taught. Data analysis at multiple levels (classroom, school, district) reveals system alignment concerns.

Test Development

The Summit Mathematics Benchmark Assessments consist of three assessments per year for K-8 and the six courses of high school mathematics. Both the traditional and integrated high school tests assess the same standards with the same items over the course of three years. All items are multiple-choice with the exception of kindergarten where the test consists of individually administered performance tasks. The first grade test is read to students in small or large group settings.

The benchmark assessments were developed in a standards-based process involving both district and state mathematics leaders. Essential standards for each grade level and high school course were first identified using Doug Reeves' criteria for power standards (Reeves, 2002). An articulation of the essential standards ensured these standards represented a coherent development of the mathematics through K-8 and high school. The essential were not always standards that could be appropriately assessed in a multiple-choice format, but were used to inform which standards would be assessed on each benchmark. In these cases, other closely related standards were chosen. The resulting standards assessed are often foundational prerequisite skills or the culmination or application of several other grade level standards. Using OSPI's Test and Item Specifications (Office of Superintendent of Public Instruction, 2009) and item writing process, items were developed and reviewed for each identified standard to match the state summative assessment.

One purpose of the benchmark assessments is to provide teachers with information about student learning of specific state mathematics standards. In order to make decisions about the mathematics program and instruction as was important that the assessments align with the curriculum sequence, assessing standards *after* they have been taught (Perie, Marion & Gong, 2007). Therefore, some Summit districts customized pacing guides to align to the benchmark assessments. In other districts, the benchmark assessments were customized to match the pacing guides.

Positive Results

District and teacher leaders experienced several power outcomes during the implementation of the Summit and Benchmark Assessments. The implementation of the Summit Mathematics Benchmark Assessments provide district and teacher leaders the opportunity to deepen their understanding of the state mathematics standards and their assessment. After the administration of the first benchmark, districts were surprised by the level of rigor of mathematics content and vocabulary. District and teacher leaders' questions led to conversations around the mathematics content in specific standards and the vocabulary requisite to teach those standards.

For example, items on the third grade assessment used the vocabulary word *inequality*. Many questioned the appropriateness of the word *inequality* at the third grade level. After further inspection of the standard and the OSPI Test and Item Specifications, district and teacher leaders had a better understanding of the content and how it is assessed. In many instances, the benchmark assessments highlighted the importance of teachers using and expecting students to use correct mathematical vocabulary.

In other instances, district and teacher leaders found the manner in which standards were assessed on the benchmark assessment was different than students' classroom experiences. In these cases the benchmark assessments illuminated alignment concerns where instructional materials were not fully aligned with state standards and the manner in which these standards are assessed on the state summative assessments.

The most powerful outcome of the Summit Mathematics Benchmark Assessments is the collaborative dialogue around data analysis. Adapted from the Data Wise Instructional Improvement Process (Boudett, City, & Murnane, 2008), DSIA designed a process and protocol to structure district's data analysis of the benchmark assessments results. Professional development and on-site technical assistance was provided to each district on the use of the process for data analysis and decision-making. Teacher teams use benchmark assessments data tied to specific standards and the misconceptions for each distracter to define a specific student learning problem. The distracter analysis reports provide the error or misconception the student linked to each incorrect response. This report is available at the

Continued on next page

classroom, school and district level.

The data analysis process provides a structure for teachers to analyze current curriculum, instruction, and assessment practices related to a specific standard and student learning problem and plan for improvement.

These conversations deepen understanding of the mathematics in the standards and common misconceptions related to the mathematics, and help establish the fundamental link between student learning and instructional practice (Boudett, City, & Murnane, 2008). Applied to multiple levels of the system, the process can also be used to guide decision making at the school and district levels.

Problem Solving & Next Steps

Although there have been many positive results in the field testing of the benchmark assessments, there are many improvements that will need to be made to make this project both sustainable at the district level and scalable to other districts served by DSIA. Problem solving is currently occurring around two primary issues: providing results in a timely manner and improving accuracy of student roster and course data. All technology teams involved are working to improve upon the accuracy of student roster and course data so that assessment results can accurately be linked to students, classes, and schools. Work is also currently underway to ensure the design of the answer sheet minimizes scanning errors. For example, a small solution may include enlarging the font size on the answer sheet to ensure more accurate scanning. However, much larger problem solving must occur around both of these issues in order to streamline test administration and improve the ability to use data to make timely instructional and program decisions.

The information gathered in the field test, including data and item analysis, is being used to guide further development and a rigorous revision process. While many initial results in the use of the data were positive, a technical evaluation will need to occur to determine the effectiveness of the benchmark assessments. All stakeholders are working diligently to ensure the Summit Mathematics Benchmark Assessments realize their fullest potential.

References

Barr, R. & Parrett, W. (2007). *The kids left behind: Catching up the underachieving children of poverty*. Bloomington, IN: Solution Tree.

Boudett, K., City, E. & Murnane, R. (2008). *Data wise: A step-by-step guide to using assessment results to improve teaching and learning*. Cambridge, MA: Harvard Education Press.

Marshall, K. (2006). Interim assessments: Keys to successful implementation. New Leaders for New Schools. Retrieved from: http://www.marshallmemo.com.
Office of Superintendent of Public Instruction. (2009). Test and item specifications: Grades 3 – 5 mathematics. Olympia, WA: Office of Superintendent of Public Instruction.

Office of Superintendent of Public Instruction. (2009). Test and item specifications: Grades 6 – 8 mathematics. Olympia, WA: Office of Superintendent of Public Instruction.

Perie, M., Marion, S. & Gong, B. (2007). A framework for considering interim assessments. National Center for the Improvement of Educational Assessment. Retrieved from: http://www.nciea.org/publications/ConsideringInterimAssess_MAP07.pdf.

Reeves, D. (2002). *The leader's guide to standards: A blueprint for educational equity and excellence*. San Francisco: John Wiley and Sons.

Schmidt, W.H. & Cogan, L.S. (2009, November). The myth of equal content. *Educational Leadership*, 67 (3), 44-47.

--Shannon Edwards is Mathematics Director for OSPI in the District and School Improvement Accountability Tacoma office. She was a former math instructional coach at Chief Leschi Schools and math teacher for Franklin Pierce School District. Contact her at shannon.edwards@k12.wa.us .

Accountability Index Goes Live

By Pete Bylsma

At its January 2010 meeting, the State Board of Education (SBE) approved a new method for measuring school and district effectiveness. The creation of an Accountability Index was a culmination of years of work by SBE after it took over the accountability duties of the A+ Commission.

The Legislature initially required SBE to "adopt objective, systematic criteria" to identify schools and districts for *recognition* and *further support*. ESHB 2261, passed during the 2009 legislative session, included more details about accountability and required the creation of the Index.

Figure 1
The 20-cell accountability matrix

	OUTCOMES				
INDICATORS	Reading	Writing	Math	Science	Grad Rate
Ach. of non-low income students					
Ach. of low-inc. students (FRL)					
Achievement vs. "peers"					
Improvement from previous year					

How the Index is Calculated

The Index is based on how schools and districts perform on a set of five outcomes and four indicators. The five outcomes are the state test results in four subjects (reading, writing, mathematics, science) in all the tested grades and the "extended" graduation rate for high schools and districts. These five outcomes are examined in four ways (indicators):

(1) achievement of non-low income students;
(2) achievement of low income students (percent meeting standard);
(3) achievement of all students compared to similar "peer" schools/districts (controlling for levels of ELL, special education, income, mobility and gifted); and
(4) improvement on the Learning Index from the previous year.

This creates the 20-cell matrix shown in Figure 1.

Each cell of the matrix is rated on a 7-point scale based on a set of fixed criteria. All subjects have the same set of benchmarks, and the assessment results are the aggregate totals for all the tested grades (results are not generated by grade or grade band).

Districts and schools have the same rating system, and fixed criteria makes it easier to understand the rating system.

For example, the ratings in Figure 2 apply to both the non-low income and low income achievement cells in all four tested subjects. (Achievement is measured in terms of the percentage of students meeting standard.)

Figure 2
Index ratings as a function of percent meeting standard

90-100% 7
80-89.9% 6
70-79.9% 5
60-69.9% 4
50-59.9%.....3
40-49.9%.....2
< 40% 1

The "peers" and "improvement" indicators are measured using the Learning Index. This focuses attention on all students across the achievement spectrum. As a result, there is no longer an incentive to focus on students who are close to meeting standard ("on the bubble").

The Accountability Index is the simple average of all the rated cells. Districts and high schools have 20 cells, while the typical elementary and middle school has 16 cells (there are no ratings for graduation). Disaggregated results for all the subgroups (e.g., race/ethnicity, ELL, special education) will still be reported as required by NCLB, but they are not used to calculate the Index. (Subgroup accountability will be address through a separate matrix of indicators and outcomes.)

Results for ELLs are not included in the first 3 years of enrollment in a U.S. public school. To increase accountability for these students, WLPT results will be posted on the Web site, and OSPI plans to report on how many ELLs are "on track" to meeting standard based on their WLPT scale scores.

Schools and districts are placed in one of five "tiers" based on their index. The table below shows the tier names and index ranges. The table also shows the **results for 2009**. These results show that the Index sets a high bar for performance – only 43% of the schools had an Index that was in the Good tier or better.

Continued on next page

Table 1
Tiers for each index range in 2009

Tier	Index Range	Pct. of schools	Pct. of students
Exemplary	5.50 – 7.00	5.0%	3.7%
Very Good	5.00 – 5.49	6.6%	6.3%
Good	4.00 – 4.99	31.0%	34.1%
Adequate	2.50 – 3.99	45.4%	48.9%
Struggling*	1.00 – 2.49	12.1%	7.0%

* 247 schools were in the Struggling tier and enrolled 71,000 students (1 in every 14 statewide); 96 were alternative schools or served other special populations.

OSPI and SBE are working together to see if the Index can be used for federal accountability purposes in light of its many advantages. Besides being a more accurate reflection of school performance, the Index is more inclusive and rigorous than the federal AYP system. For example, it includes both writing and science, uses a smaller minimum number for reporting (10 students across the entire school/district), and uses results of all students regardless of how long they attended school. Moreover, no margin of error is used to adjust the results.

The Index is also fairer than the current AYP system. Averaging all the rated cells means the system is "compensatory" in nature—one low rating does not automatically result in a school/district not making AYP. Using a compensatory approach helps us maintain high performance standards. Two studies found that Washington has some of the nation's toughest AYP requirements, resulting in a high number of schools "needing improvement."[1] Using the Index reduces the incentive to lower our content and performance standards so all students can be proficient and meet federal targets by 2014.

Using the Index for Recognition

The Index cannot be used for AYP purposes this year. However, it is now being used for recognition purposes. In March, the joint OSPI/SBE **Washington Achievement Award** was developed and used the Index in a norm- and criterion-based way. In total, six awards were created.

The Outstanding Overall Performance award is for schools whose overall 2-year Index average puts them in the top 5% in their grade band: elementary, middle/junior, high and multiple grades.[1] Schools must have had at least 10 cells of the matrix rated each year and fewer than 10% students designated as gifted each year to be considered.

Table 2 shows how many schools of each grade band received this type of recognition.

Special Recognition awards are given to schools that are top performers in five areas:

1. Language arts (reading and writing combined)
2. Math
3. Science
4. Extended Graduation rate
5. Gifted Education

In the first four award areas, the 2-year "column" average must be at least 6.00, at least 2 of the 4 cells in the column must be rated each year, and there must be fewer than 10% students designated as gifted each year. For gifted education, recognition is given to schools that have at least 10% gifted students in both years, a 2-year peer "row" average of at least 6.00, and at least 2 cells are rated in the peer row each year. Table 3 (right) shows now many schools were recognized for these types of awards.

Table 2
Number SY 2008-09 awards to schools in SY 2009-10 by index

Grade Band	# in top 5%	Index cut-off	Total awards
Elementary	53	5.280	70
Middle	19	4.875	26
High	20	4.910	52
Multiple	16	4.735	26
Total	108		174

Table 3
Numbers of SY 2008-09 schools for special recognition

Recognition Categories	Numbers of schools recognized
Language Arts	36
Mathematics	10
Science	24
Graduation Rate	35
Gifted	20
Total	125

[1] See "*The Accountability Illusion*," Thomas Fordham Foundation (February 2009) and "*Schools Struggling to Meet Key Goal on Accountability*," Education Week (January 7, 2009).

Continued on next page

In total, 174 schools were recognized in 233 areas, and 48 schools received recognition in more than one category. Other award categories will likely be developed in the future.

Two years of Index data have been posted on OSPI's Accountability Web site (www.k12.wa.us/Accountability). Index data are now being used by educators around the state to help identify areas of greatest need. More details about the system can be found on this Web site.

--Pete Bylsma is an independent education consultant to the State Board of Education and is a past WERA President. Contact Pete at bylsmapj@comcast.net.

[1] For the 2009 award, the "2-year average" was based on the average of data from 2008 and 2009. The top 5% is based on the total schools at that level in the 2009 index (this includes schools that did not receive an index score). When a tie occurs at the 5% cut off, all schools with that score are recognized.

Introducing Dr. D'LEma

A WERA Journal Forum on Data and Research Ethics

So who is this illustrious Dr. D'Lema, anyway? We are all Dr. D'Lema now.

We live in an increasingly complex world of data and information, as well as *mis-data* and misinformation. Have you noticed that our roles and responsibilities as data professionals have changed dramatically in recent years ? Perhaps you, too, have encountered ethical dilemmas in your daily work as you collect, analyze and report data about students, educational programs, schools and school districts. Where to begin?

> "Although the world is full of good people, ethical behavior does not just happen automatically …What is 'common sense' to one person is not necessarily to another (*NFES Code of Data Ethics*, 2010, p. 2- 3).

This column is intended to give voice to some of these concerns and provide a meeting place for dialogue and discussion, out of harm's way. Where might this effort lead? Intended outcomes include, but are not limited to:

- A Data Ethics Advice Column, inspired by Randy Cohen's *Ethicist* Column, featured in the *New York Times*, and several other newspapers;
- A place to share book reviews, on-line publications, scholarly articles, and other types of discourse relating to data, assessment, program evaluation and research ethics;
- A forum to discuss local news, as well as workplace issues, relating to data, assessment, program evaluation and research ethics;
- Advocacy – taking a stand(s) or advocating a position(s) related to education, assessment, program evaluation and data ethics.

We look forward to a spirited discussion and encourage you as WERA members and citizens of the education community to join us. We encourage you to submit original articles, as well reviews, letters to the editor, and other thoughts and commentary.

To launch this new enterprise, we start with a review of the *National Forum on Educational Statistics Code of Data Ethics (2010,)* just released, by Andrea Meld, data professional at OSPI and past WERA Board Member, mindful that the *NFES Code of Ethics* will be a valuable foundation for further exchange.

--Dr. D'Lema may be reached by addressing inquiries to the editor at p.hendrickson43@gmail.com
He will forward to D'Lema.

Do We Need a Code of Data Ethics? A Review of the National Forum on Educational Statistics *Code of Data Ethics*

By Andrea Meld, Ph.D.

What are Data Ethics?

Conduct a Google search on "Data Ethics" and you are likely to find over 37 million results. Clearly this is not an esoteric topic. Professional organizations such as AERA, APA, and others have worked hard to establish and publish guidelines for fair testing practices, determine standards for the use of human subjects in experimentation, and there are codes of ethical conduct for those in the medical and mental health fields. So what are data ethics? Are we talking about the avoidance of harm or the impression of wrong-doing or simply obeying a set of rules? Doesn't the Family Educational Rights Privacy Act (FERPA) cover this already? In what ways does ethical reasoning (Kohlberg, 1971) and action differ from simply following the legal codes or administrative guidelines?

In the introduction the authors state:

Each and every day, educators collect and use data about students, staff, and schools. Some of these data originate in individual student and staff records that are confidential or otherwise sensitive. And even those data that are a matter of public record, such as aggregate school enrollment, need to be accessed, presented, and used in an ethically responsible manner. While laws set the legal parameters that govern data use, ethics establish fundamental principles of right and wrong that are critical to the appropriate management and use of education data in the technology age. (Forum Guide to Data Ethics, 2010, p. iv)

The *Code* distinguish between laws and ethics. While laws may set the legal parameters for data users to operate…ethics go deeper and are often more rigorous. For example, it is one thing to change the scale on a graph, but another to deliberately mislead the reader.

Quite apart from a data professional's rank or role in an organization, the consistent demonstration of honesty, integrity, and professionalism are of greatest importance. "These qualities, more than any other characteristic or trait, serve as the foundation of ethical behavior." On when to address important issues, Rabbi Hillel said, "If I am not for myself, who will be for me? And if I am only for myself, then what am 'I'? And if not now, when? "Say not, 'When I have free time I shall study'; for you may perhaps never have any free time." (Pirkei Avot, 2010).

Background and Purpose of The Forum Guide to Data Ethics

Education organizations need a simple, comprehensive set of standards for establishing plans that encourage the ethical use and management of data. In response to this need, this document presents a code of ethics for data management and use in education settings.

While the Code states that ultimate responsibility for data ethics rests with leadership, all others with access to student and education data are charged with understanding and following the appropriate ethical behavior in accessing, using, and managing education data. Named are superintendents, chief information officers, principals, teachers, registrars, counselors, school board members, data managers, technology directors, information systems staff, data stewards, technical staff, and other data professionals and office staff. (Sounds like the WERA mailing list).

This approach brings to mind what Lawrence Kohlberg (1971) referred to as post-conventional morality. The late moral develop mentalist's view was that rules, while generally useful, may and sometimes *must* be changed to protect human rights and values. Rules are not absolute dictates to be obeyed without question, he wrote.

Core Principles of the Code of Data Ethics

At the heart of the Code is an Integrity Canon of core principals, brief statements we could stare at each day as we deal with sensitive data.

The Integrity Canon

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

The Data Quality Canon

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

The Security Canon

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

It is instructive to compare these core principles to advice from *Pirkei Avot* (2010), "Pirkei_Avot&action=edit&section=7 Show kindness to others, respect the other person, Pirkei_Avot&action=edit&section=9 respect yourself,

Continued on next page

Pirkei_Avot&action=edit&section=13 be humble, Pirkei_Avot&action=edit&section=14 immerse yourself in learning, Pirkei_Avot&action=edit&section=18 be careful with speech, Pirkei_Avot&action=edit&section=20 do not leap to judge another person, Pirkei_Avot&action=edit&section=21 be fair and deliberate, Pirkei_Avot&action=edit&section=23 the time for action is now." For an even 10, I submit another canon: Consider the unintended consequences and consequential validity of data collection and reporting practices for your community and organization (Messick, 1989).

Vignettes Drawn from Real Life

For each canon, a real-life vignette is taken from a real-life situation to illustrate each core principle. These may seem familiar to you. Following is an example of a vignette designed for discussion of data integrity:

The community was thrilled to learn that the local high school had been named one of the top 10 schools in the country by a major news magazine. However, when examining the methodology behind the award, the district superintendent questioned the finding and decided that she needed to know how the rankings were determined. An inquiry to the magazine found that the data had been "checked and double-checked," but no one at the publication was willing to divulge what data were used to determine the rankings. Additional investigation by district staff revealed that the magazine had used an incorrect enrollment figure, causing the participation percentage on a national test to be tremendously inflated. The superintendent understood that, if she reported this to the magazine, the high school would surely drop from the top tier to the second tier of "best schools." Still, the error had to be corrected—it was the right thing to do. Despite the decline in national prominence, the superintendent was surprised to learn that her community—including parents, students, alumni, and the local media—were very proud that the school district chose to report the error rather than receive recognition it didn't deserve. Ensuring accuracy over fame had actually confirmed to community members that they really did have one of the top school systems in the country. (Forum Guide to Data Ethics, 2010. p.15)

The authors point out that ethics are sometimes put to the test when data show something unexpected or negative, and the need to impartially report bad as well as good news.

Examples of Recommendations for Professional and Organizational Development

Each section in the canon provides a list of recommended practices and training for those in the chain of data usage. Below is an abbreviated synopsis.

- Create an organizational culture that encourages honesty, integrity, and professionalism.
- Before releasing a report, undertake an independent review to assess whether the data are presented objectively and without bias, especially when they describe a situation that is not favorable to those responsible for producing the report.
- Be accountable. And hold others accountable for ethical use of data so that they may report suspected violations without fear.
- Be very cautious about using data for purposes other than their original intent. Be sure that doing so does not violate individuals' right to privacy or any agreements of anonymity that you, or your agency, have made.
- Accept that there are limits to how well data can describe people—*e.g.* people with complex thoughts, needs, and emotions; people with physical or psychological challenges that may not be well understood; or people who, through no fault of their own, live in circumstances that are unhealthy, unsafe, or unstable. Each piece of data in a student or staff database represents an attribute of a real persons, but these data cannot adequately portray all aspects of multifaceted individuals.
- Be especially careful about making personal or professional judgments about people based solely upon data. Be particularly alert to data that may be flawed, narrow in scope, or otherwise of limited applicability.
- Effective, data-driven decision-making draws from multiple sets of data that support the same interpretation. Do not make decisions based upon a single source, if at all possible. Review data from multiple sources over time to see if the findings are consistent.

Self-Assessment

To test your own understanding of the *Code of Data Ethics*, see if you can match each direct quotation below to the appropriate canon. At a more complex level, discuss these statements with a colleague(s) at work to see if you agree or disagree, and upon what basis:

1) Just because data can be used to answer a question or inform an opinion does not mean that the information is entirely accurate, reliable, and unbiased.
2) Be willing to challenge commonly held assumptions and prejudices related to descriptive data. For example, do not equate disability status with decreased intellectual aptitude or potential.
3) Staff who consistently demonstrate honesty, integrity, and professionalism are the foundation of ethical behavior in an education organization.
4) Avoid at all costs the release of any data that could lead to physical, mental, or emotional harm to others.
5) Data should provide information that is useful to the organization in a practical way. If data are not useful, there is no reason to collect them.

Continued on next page

6) Say what you mean, and mean what you say. For example, deceiving respondents by implying that you are collecting data for the district when it is really for a master's thesis is ethically untenable under all but the rarest of circumstances.

Answers and commentary will be posted in the next edition of the WERA Journal.

A Final Note

The authors remind us that school is school, only one part of complex lives, "Do not automatically equate school success with life success. Academic success is important, especially within the context of the education system, but people can find happiness, prosperity, and success in life without being the highest achiever in school (Forum Guide to Data Ethics, 2010, p. 12)."

References

Ethics of the Fathers/Pirkei Avot. (2010). *Jewish Virtual Library*. Retrieved from http://www.jewishvirtuallibrary.org/jsource/Judaism/pirkei_avot.html. *(Note: The Pirkei Avot with Hillel's influential teachings was published circa 200BCE-200CE.)*

Kohlberg, L. (1971). *From* Is *to* Ought*: How to Commit the Naturalistic Fallacy and Get Away with It in the Study of Moral Development*. New York: Academic Press.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.) New York: Macmillan.

National Forum on Education Statistics. (2010). *Forum Guide to Data Ethics (NFES 2010–801)*. U.S. Department of Education. Washington, DC: National Center for Education Statistics.

NCES World Wide Web Electronic Catalog -- *http://nces.ed.gov/pubsearch*

--Andrea Meld is an OSPI data analyst and editor of WERA's The Standard Deviation. *You may contact her at andrea_meld@g.com*.

Opinion: Textbook Selection and The Role of Research

By Peter Hendrickson, Ph.D.

For the first time in my education career in four countries, three states, and several districts, textbook selection in my local community is being challenged in the courts (Long, 2010). At the same moment federal initiatives are calling for bold, provocative, even disruptive, changes to schools and districts to break patterns of persistently low student achievement (Duncan, 2009). The call for evidence-based instruction seems to have passed by the practice of selecting texts for use in student classrooms. Witness the culture wars playing out in Texas around the content of history texts, stark evidence that schools are rarely an apolitical arena.

Unlike the few states that provide vetted lists of approved textbooks (CA, FL, IN, KY, NC, OR, TX…), Washington has only recently provided recommended math and reading textbooks. Supt. Randy Dorn issued a revised list of recommended and not recommended mathematics textbooks in 2009 (Dorn, 2009). Two texts were recommended for elementary, three for middle grades and a single text for high school mathematics. Several other texts were explicitly not recommended based on reviews by consultants from outside Washington. As districts do not receive explicit funding for texts linked to the recommendations, the Memoranda do not have the same force as those selections in the states where districts may only purchase from the state lists to qualify for funding.

In February Seattle Schools lost a lawsuit over their selection of *The Discovering Mathematics* series, which was reviewed by not recommended by OSPI (Long, 2010). Competitor *Holt Mathematics* was up for consideration against *Discovering* in both Bellevue and Issaquah. Bellevue last fall assigned competing texts to different high school classrooms. Those classrooms were matched on demographic and test score variables but texts were not randomly assigned. Outcome measures were student grades over selected units on common classroom assessments. A district curriculum committee weighed the results and recommended *Holt* which the Board approved in April. In Issaquah *Discovering* was recommended.

Figure 1
Head to head comparison of Holt and Discovering series in Bellevue schools
Data Source: Save Issaquah Schools (2010)

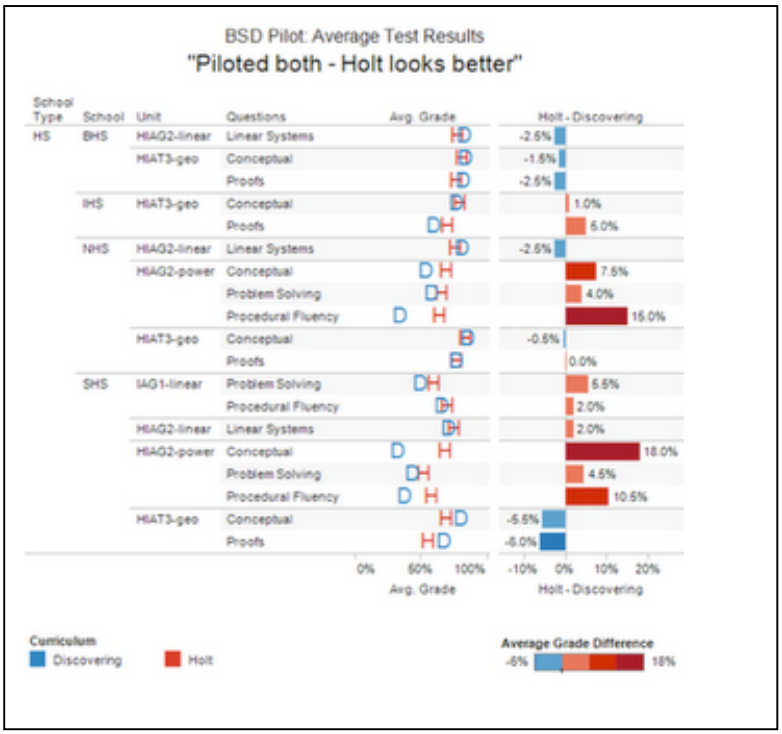


Figure 1 shows the unit test score comparisons--percent correct-- from Bellevue where the texts were assigned to different classrooms across high schools. Longer red bars show a greater difference in student scores favoring the *Holt* text.

The primary charge to the OSPI consultants who analyzed texts offered for gaining recommended status was to match the textbook content and methodologies with (changing) state standards. One evaluator noted that while a careful read of the texts could yield a degree of match to standards, there was no investigation of the student achievement outcomes to compare program effectiveness.

An international review of textbook research (Nicholls, 2003) across several nations noted that is essential to see how texts are used by teachers and

Continued on next page

received by students in addition to the content of the textbook. The British study also noted that we continue to know very little about the impact of using textbooks.

Even the American Association for the Advancement of Science (AAAS) math and science Project 2061 textbook reviews were conducted by distinguished panels of experts examining weighty books, not by reviewing impact studies published in Tier 1 academic journals (AAAS, 2010).

Tier 1 research with random assignment to treatment (different textbooks) and careful selection of the sample (different classrooms) is exceedingly difficult to perform in our schools. Bellevue's effort to conduct quality research is commendable but most textbook decisions will be determined in complex political (and more often polarized) setting, rather than one buttressed by the knowledge that students will learn more of what we wish them to learn with this text rather than that text. Perhaps the federal Department of Education could be persuaded to support bold, even disruptive, research to examine the impacts of various textbooks with highest quality research designs. They might avoid the prevailing culture wars if they remained in the science, technology, math and reading arenas—but likely not.

References

AAAS. (2010). *Project 2061 textbook evaluations*. American Association for the Advancement of Science. Retrieved March 23, 2010 from http://www.project2061.org/publications/textbook/default.htm.

Dorn, R. (2009). *Memoranda MO 014-09M, MO28-09*. Olympia, WA: Office of State Superintendent of Public Instruction. Retrieved from http://soundmath.wetpaint.com/ March 2, 2010.

Duncan, A. (2009). Talk given to the Brookings Institution May 11, 2009. Access transcript at http://www.brookings.edu/~/media/Files/events/2009/0511_duncan/20090511_education.pdf

Long, K. (2010 a). Textbook decisions making districts wary. *Seattle Times*, Section B, p.1., February 13, 2010.

Long, K. (2010 b). Q & A on math textbook controversy. *Seattle Times*, Section C, p.1, February 24, 2010.

Nichols, J. (2003). Methods in school textbook research. *International Journal of Historical Learning, Teaching and Research*, V3, N2. Published online at http://www.ex.ac.uk/education/historyresource/journalstart.htm.

Save Issaquah Math (2010). retrieved from http://saveissaquahmath.blogspot.com/2010/03/march-18-2010-to-issaquah-school.html.

--Peter Hendrickson was a curriculum director in Ohio, Alaska and Washington school districts prior to specializing in assessment, research and program evaluation. The past WERA president is now an OSPI program evaluation contractor with concurrent evaluation work in traumatic brain injury research.

WERA Book Reviews

In this issue we share three book reviews, all relating to different aspects of data. As practitioners look for more sophisticated ways to look at achievement differences across individuals and groups, these books seem particularly timely.

Andrea Meld revisits a classic in this area with her review of Singer and Willet's *Applied Longitudinal Analysis*. **Bob Isenberg's** review of Few's *Information Dashboard Design* focuses our attention upon another challenge of the data practitioner: displaying data in a way that is both accessible and illuminating for a variety of consumers. Finally, **Michael Power** moves us into the arena of the multiple interpretations of data with his review of Michaels' *The Trouble with Diversity: How We Learned to Love Identity and Ignore Inequality*.

These three reviews definitely offer something for everyone concerned with the uses of data. For even more stimulation, you are directed to Andrea Meld's review in this issue of *Code of Data Ethics* from the National Forum on Educational Statistics.

--Phil Dommes, Book Review Editor, North Thurston Public Schools

Book Review: *The Trouble with Diversity: How we learn to Love Identity and Ignore Inequality*

By Walter Benn Michaels

Reviewed by Michael Power, Ph.D.

Walter Benn Michaels has a problem with our society's obsession with celebrating diversity. He thinks it's a scam.

Benn Michaels blames both the left and the right of the political spectrum in the US, and charges them with being in accidental collusion on distracting society from the real problem – massive inequities in economic status. We focus upon the celebration of years spent overcoming racial and ethnic discrimination rather than tackling these persistent economic inequalities. He cites universities (Benn Michaels is a Professor of English at the University of Illinois at Chicago) as particularly active in this area – touting their highly diverse student bodies while ignoring that many of these students of color are economically indistinguishable from their white classmates.

Why celebrate diversity at the expense of social justice? Benn Michaels proposes that it is due to our collective unwillingness to take on the huge challenge of correcting generational poverty. This would require a shift in economic resources unlike anything in our history, and would upset a number of very comfortable apple carts – both liberal and conservative.

Ralph Nader often remarks that the "left" and "right" in US politics are as different as Tweedle Dum and Tweedle Dee. In terms of commitment to and policies of social justice, Benn Michaels would agree.

> ". . . it testifies to the triumph of racial prejudice over class privilege, which is to say, the way in which it demonstrates the irrelevance of wealth and (from the standpoint of the racist) turns class warfare into white supremacism while (from the standpoint of the antiracist) turning class warfare into bigotry. If you're a racist, it shows you that racism is the solution; if you're an antiracist, it shows you that racism is the problem. Either way (racism) is a kind of gift since it makes over the rational anger of the poor as the irrational anger of the racist and enables everyone to agree that the real issue here is not money but race" (p. 67).

What would trouble some readers, even those in general agreement with his thesis, is Benn Michael's willingness to take bits of emerging scientific evidence and generalize from them as if they were common knowledge and are impacting our behavior and attitudes on a day-to-day basis. For example, in Chapter 1 (*The Trouble with Race*), he jumps from the recent findings from DNA research showing that the variability within races is at least equal to the variability among races. He then extends this finding to the proposition underlying the rest of the text--that there are no biological differences attributable to race. While this may be a justifiable conclusion on a biological level, the reality is that people generalize based on groups, not individuals. As a group, there are substantial differences between Caucasian Norwegians and Black Nigerians (Benn Michaels' examples of in-group variability) in terms of how they look and their cultural practices. Prejudices and discrimination based upon race continue to be a major problem in the US and elsewhere, whether they are justified by biological data or not.

This does not diminish Benn Michael's larger argument that we need to tackle poverty head on and stop patting ourselves on the back for the "progress" which doesn't make a real difference in resolving economic inequity. However, it does nag at the reader every time he states a variant of "Why does racial difference remain so important to us when the racism it was used to justify is so widely condemned and when the basic idea about race that gave it its power – the idea that there are fundamental physical or cultural differences between people that line us with our division of them into black, white, et cetera – has been discredited?" (p. 49). The correlation between economic status and race is powerful. Many other commentators have attributed this not only to an unwillingness in our society to address the economic issues, but also to systemic racism independent of affluence.

Benn Michaels concludes with a powerful indictment of our current political system.

> "People sometimes say that economic and ideological issues put voters off. But almost half the population already doesn't vote. And most of the nonvoters are poor. It would be interesting to see how many of them would start showing up for elections that hinged on the meaning and importance of equality in America" (p. 190).

As the Obama administration shifts the conversation at the national level toward addressing issues of poverty through education and community development (e.g. the upcoming US DOE Promise Neighborhoods grants and the upcoming Housing and Urban Development Choice Neighborhoods grants), we may have a chance to answer Benn Michael's question.

Publication Data: *The Trouble with Diversity: How We Learned to Love Identity and Ignore Inequality* by Walter Benn Michaels, 2006. Henry Holt and Company, New York, NY, Paperback. 241 pages, $15 (US) ISBN-13: 978-0-8050-8331-6

--Michael Power is Manager of Educational Programs for the Tacoma Housing Authority and a past-president of WERA. Contact Michael at mpower@tacomahousing.org

Book Review: *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*

By Judith D. Singer and John B. Willett

Reviewed by Andrea Meld, Ph.D.

We are in the midst of another paradigm shift in our work with education and assessment data analysis. Many people, myself included, who worked with static, cross-sections of student data, linking each successive wave of reading and math scores for seventh graders, for example, looked forward to the day when we could track cohorts, describe individual student growth, and focus on the differential rates of educational achievement characteristic of the achievement gap, with anticipation of research well devised, implemented, and received.

This approach, intuitively, seems like a much better way to measure growth and change over time. But what do we really know about longitudinal data analysis and research in practice? What do we need to learn? Many "newer" approaches, such as HLM may have started at a time when we were long past graduate seminars in educational measurement, and were launching our professional careers.

Although it was published in 2003, *Applied Longitudinal Data Analysis* fills a gap between the requirements of new educational assessment policies and directives and our understanding of how to conduct and apply longitudinal analysis. I had heard that this book was considered the "Bible" of longitudinal analysis. It is well-written and may inspire humility or zealotry among new converts. It reminds the reader to analyze data wisely and with caution, sometimes with wonder.

Stephen W. Raudenbush, The University of Michigan, remarked that *Applied Longitudinal Data Analysis* emphasizes that statistical models are only "tentative representations, subject to criticism and revision based on data. It wages a much-needed struggle against overly formulaic thinking that is all too common in the everyday practice of statistical analysis in social science." (See: Author Information link)

The authors begin with a definition of change and two types of questions that are fundamental to every change study. First, what is the pattern of an individual's change over time? Is it linear? Does it fluctuate? The second question invites the reader to compare individual's in terms of change and any predictor variables.

I should relate this to an example from education, but I keep thinking of diets we've tried over the years. Was the pattern of weight loss consistent or did it fluctuate? Linear or non-linear? Continuous or did it stop? How can different people on the exact same diet plan lose, stay the same or even gain weight?

According to the book, three important features for a study of change are: 1) at least three waves of data; 2) an outcome that changes systematically over time; and (3) a sensible way to measure time. The authors explain these details with clarity and great examples.

Next they outline, illustrate, and explain how to set up a longitudinal data set. For me, this was a revelation. In fact, they make a strong case for using person-period data rather than person-level data. Person-level, cross-sectional, or multivariate data is probably the arrangement we are most familiar with and is widely used. It is characterized by one single row for each subject or record, regardless of wave, with the variables to be measured in columns or fields. The main advantage, according to the authors, is that you can visually trace growth from older data to the left towards newer data on the right. However, there are many disadvantages of person-level data for longitudinal analysis. In sum: 1) summaries are non-informative; 2) no explicit tie to the variable; 3) the number and spacing of waves that vary across individuals; and 4) the exclusion of time-varying predictors. Much greater detail is provided in this section.

By contrast, in the person-period, or univariate data set, each person's empirical growth data is shown vertically rather than horizontally. There will be fewer columns of data, but many more rows. The person-period data set consists of four types of variables: 1) a subject identifier, 2) a time indicator, 3) outcome variable(s), and 4) predictor variable(s). Storing data in this format makes it much easier to accommodate any data collection schedules and any combination of time-varying and time invariant predictors. If you are not convinced, the book features data arrays and makes a persuasive case for this way to structure longitudinal data. If you have ever worked with this type of data using the person-level approach, you may need no convincing.

The next two chapters introduce and describe data analysis with the multilevel model of change. But I think I will stop here. Much of the book is available on-line. This topic is timely with Race to the Top and other new proposals for measuring student growth and closing the achievement gap.

About the Authors: Judith D. Singer, Ph.D., is James Bryant Conant Professor at Harvard University, Graduate School of Education. John B. Willett, Ph.D., is Charles William Eliot Professor, Harvard University, Graduate School of Education. As colleagues and frequent collaborators, the order of their names in random.

Continued on next page

Publication Details: *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*, by Judith D. Singer and John B. Willett, 2003. Oxford University Press, Hard Cover, 672 pages, $74.95 list price – $55.00 if ordered directly from publisher, for a limited time. Also available as Kindle book. **ISBN-10:** 0195152964, **ISBN-13:** 978-0195152968

Data files used in the book may be downloaded in SPSS, SAS, HLM, and other packages from UCLA Technical Services ALDA support website, http://www.ats.ucla.edu/stat/examples/alda/.

--*Andrea Meld is a past WERA Executive Board member and edits* The Standard Deviation. *She has served as Data Analyst, Assessment & Student Information, at OSPI since 2003.*

Book Review: *Information Dashboard Design*

By Stephen Few

Reviewed by Robert Isenberg

The mantra of "data-based decision-making" requires getting data to education decision-makers who, whether administrators or teachers, were neither prepared to nor inclined to analyze data. Admit it; most principals' and teachers' eyes glaze over at the sight of an Excel spreadsheet. Mention of "stanine" or "NCE" and you are likely to induce a catatonic state. What to do? Increasingly, vendors and data-savvy educators tout data dashboards as a vehicle for getting the right data to the right people. Think car dashboards. For teachers, the dashboard may include assessment, attendance, and grade data about the students in their classes; for administrators, summary data of performance on state assessments by AYP cells, attendance rates, and semester grades by course. The Washington State Information Processing Cooperative (WSIPC) is developing such a dashboard. Vendors are ubiquitous.

Stephen Few's book is an excellent guide to the design of such dashboards. Though his audience is the broader business IT professional, it is a rich guide for WERA members and district IT staff. He believes that dashboards fail to deliver on their promise to provide quickly usable data. Typically there is a failure to consider overall design and user visual perception. For example, a car's dashboard is designed to give the driver what he/she needs to know now. It could provide lots more information but, then what the driver really needs to know would then be lost in the clutter. Simplicity is critical – an engine warning light appears rather than a detailed breakdown. The need to know is fundamental. The engine warning light does not stay on as green when there is no problem. No light means no immediate need to know.

Appropriately, *Information Dashboard Design* is abundantly illustrated with dashboards and specific dashboard components such as graphs. Few provides a multiplicity of visuals that reinforce his guidelines on what works, what doesn't , and why.

A flip through the book and a scan of the Table of Contents will suffice to convince you to add this book to your professional collection.

- Chapter 3: Thirteen Common Mistakes in Dashboard Design (e.g., Displaying Excessive Detail or Precision, Misusing or Overusing Color, Using Poorly Designed Display Media {one of my pet peeves – use of 3-D graphs that are pretty but hard to read})
- Chapter 4: Tapping Into the Power of Visual Perception (e.g., Understanding the Limits of Short-Term Memory)
- Chapter 7: Designing Dashboards for Usability (e.g., section on drilling down)
- Chapter 8: Putting It All Together (samples with critiques)

In all, *Information Dashboard Design* provides interesting reading and a useful reference in either the selection of a product or the creation of your own. Convince the head of your IT department to buy it and then borrow it!

Publication Information: *Information Dashboard Design* by Stephen Few**,** 2006**,** O'Reilly Media, Sebastopol, CA, Paperback, 224 pages, $34.99 (US) ISBN: 0-596-10019-7

--Bob Isenberg is the Director of Assessment with the Kent School District. Contact him at bob.isenberg@kent.k12.wa.us

A Checklist for Research in Progress: An AERA Division D Tool

AERA Division D conducted an In-Progress Research Gala during the 2010 Annual Meeting in Denver, CO, April 30 – May 4. Graduate students shared their research in progress (but not thesis/dissertation research) with an audience of interested researchers, including a panel of jurors who provided feedback on the research. In-progress work included research proposals, literature reviews, conceptual-theoretical pieces, and research projects not completed.

While they were about measurement and research methodology or an application of advanced measurement and research methodology, WERA *The Standard Deviation* Editor Andrea Meld thought school researchers might find the indicators useful. The checklist was prepared for a similar function in 2009. Organizers for 2009 were Dubravka Svetina (Arizona State Univ) dsvetina@asu.edu and Jill Adelson, then at University of Connecticut and now at Louisville jill.adelson@louisville.edu.
--Editor

Quality Indicators Checklist for In-Progress Research Manuscripts

Please evaluate the above manuscript with respect to each of the following categories by checking ONE box. Please provide detailed comments to the graduate student on what s/he did well and what s/he can improve.

(0 = No Evidence, 1 = Novice, 2 = Emerging, 3 = Skillful, 4 = Accomplished)

1. Conceptual Framework, Background, and Research Question(s)					
	0	***1***	***2***	***3***	***4***
A. A clear statement of the problem is presented, with important and relevant concepts defined.	☐	☐	☐	☐	☐
B. A clear link is established between what is known (existing literature) and what needs to be known, exhibiting an adequate grounding in theory.	☐	☐	☐	☐	☐
C. The research question(s) are clear and answerable.	☐	☐	☐	☐	☐
D. The questions posed can feasibly be addressed in the context of this manuscript.	☐	☐	☐	☐	☐
E. Posing and addressing the question(s) can be done in an ethical manner.	☐	☐	☐	☐	☐
Comments:					

2. Methodology					
	0	***1***	***2***	***3***	***4***
A. The research design is described in detail, with particular attention to variables/concepts to be addressed.	☐	☐	☐	☐	☐
B. The sample and/or sampling plan are described and are appropriate.	☐	☐	☐	☐	☐
C. Procedures for data collection are described in detail and are appropriate to the questions posed/ In simulation studies, programming plans/procedures are addressed.	☐	☐	☐	☐	☐
D. Data collection tools (e.g., instruments, interview or observation guides) are described in detail and are appropriate (if applicable).	☐	☐	☐	☐	☐
E. Data analysis plans are appropriate and are described in detail, with particular reference to the research question(s) posed.	☐	☐	☐	☐	☐
F. Figure or conceptual model is appropriate and describes the planned research (if applicable).	☐	☐	☐	☐	☐
Comments:					

Continued on next page

3. Implications					
	0	***1***	***2***	***3***	***4***
A. The study's argued importance and impact are convincing.	☐	☐	☐	☐	☐
B. The study has potential to result in future research.	☐	☐	☐	☐	☐
C. The limitations of the study are acknowledged and discussed in the context of interpreting findings (e.g., generalizability, transferability).	☐	☐	☐	☐	☐
Comments:					

4. Writing					
	0	***1***	***2***	***3***	***4***
A. The proposal is written in a scholarly manner that is consistent with the research literature in the student's field of study.	☐	☐	☐	☐	☐
B. The proposal is free of grammatical, spelling, and typographical errors.	☐	☐	☐	☐	☐
C. The proposal follows a structure and organization consistent with APA or other relevant style requirements. Proper attribution is given when the work of others is used or cited.	☐	☐	☐	☐	☐
Comments:					

Please provide any further feedback for the author.

Restaurant Review: *The Rosewood Café*

Reviewed by Heather Bundeen, Ph.D.

The division of District and School Improvement and Accountability (DSIA) through OSPI has recently relocated to the Tacoma School District's Professional Development Center. With the approaching launch of the Washington Improvement and Implementation Network (WIIN) that will offer professional development and technical assistance to invited schools and districts, it seemed to be time for a local restaurant review. The Rosewood Café is the perfect place for lunch or dinner.

The Rosewood Café is not a place that you would likely stumble upon during a quick trip to Tacoma. Nestled within a North End neighborhood just minutes from the new DSIA office, this cozy eatery can provide a pleasant escape from a hectic day. As you walk in the front door, it immediately feels as though you have been invited into a home.

The Rosewood Café is the kind of place that you can easily dine alone or with a small group. The dining room features scattered wooden tables set with fresh flowers. A counter lines the front window - complete with an array of magazines and the local newspaper to keep you company if you decide to linger. Two walls consist of a series of windows that invite the daytime sunshine and provide a dark contrast for strings of fairy lights as evening falls. Any remaining wall space in the café features the work of local artists.

A reasonably-priced menu includes lunch or dinner options complete with a selection of gourmet sodas, beer on tap, and eclectic wines. There are many unique sandwiches, soups, salads, and a range of savory dinner options. I have tried numerous items and always inadvertently mumble a yum … or two. If you happen to be a vegetarian, there are many options. On a personal note, I absolutely love the bleu cheese salad dressing, any of the daily specials, and dishes with fresh basil - like the Feta Melt.

Each day, a diverse clientele fills The Rosewood Café with a rumble of conversation and light laughter. Locals from the North End flock to this café for the brie and desserts, like the open-faced cowboy cookie. Students from the University of Puget Sound frequently wander over with their parents for homestyle entrees, like the baked macaroni-and-cheese or chicken pot pie. Families walk in from the surrounding neighborhood to see their friends and to try the plentiful children's menu options. And, couples may stop by to enjoy the inexpensive yet extensive boutique wine list.

The Rosewood Café is a former corner store and retains that great vintage charm and warmth of a locally owned place that is meticulously maintained. The owner is named Barry. It is rare to go to the café and not find him there. His excitement about the newest Gig Harbor microbrews or his latest dessert creation is quite infectious. Barry clearly loves his restaurant and is thrilled to see guests enjoying the food.

To be honest, it is difficult to leave this place without smiling as the service is exceptional. The staff members are cheerful and consistently responsive to any requests. And, it is virtually impossible to finish a meal and walk out the door without a "good-bye" and "thank you" floating behind you.

The Rosewood Café is easy to find. As you are leaving Tacoma's Professional Development Center, turn east on 26th Street toward downtown Tacoma. The café will be on your left after about 2.5 miles (or approximately eight minutes). There is always plenty of street parking.

The Rosewood Café. It just feels like home.

The Rosewood Café
3323 North 26th Street
Tacoma, WA 98407
Ph: 253-752-7999
Open seven days a week from 10:30-9:00

--Heather Bandeen has been instrumental in pulling together WIIN Center Professional services and products. She is a self-proclaimed "novice foodie". Contact her at hbandeen@gmail.com.

Washington Educational Research Association
PO Box 64489
University Place, WA 98464
www.wera-web.org

The WERA Educational Journal

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

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

Food Editor
Karen Banks, Ph.D.
Data Detectives

Ethics Editor
Andrea Meld, Ph.D.
OSPI

Humor Co-Editors
Michael Power,
Tacoma Housing Authority

Bob Silverman
Education Technology Consulting

Photo Editor
Don Schmitz
Mukilteo School District

Technical Editor
Patrick Cummings,
Tacoma School District

Editorial Assistants
Jeanne Willard &
Loretta Comfort
Everett Public Schools

Andrea Meld, OSPI

Layout Designer
Michelle Sekulich
The Boeing Company

Executive Secretary
Leonard Winchell
WERA

The WERA Educational Journal is published fall and spring as an occasional online journal. Submissions are welcomed from WERA members and others. Kindly submit articles for consideration using APA format. Copy deadlines are Octover 1, 2010 and March 4, 2011.