Using School Data

To

Clarify and Address the Problem

S.M.A.R.T. GOALS

Are goals that are:		
	S Specific	Focus on the vital areas, specific targets
	M Measurable	Multiple measures, frequent, to assess learning
	A Attainable	Motivation to strive/stretch to achieve
	R Results-oriented	Concrete benchmarks to achieve
	T Time-bound	Builds accountability and commitment
They must be:		
	• Aligned with the broader, overarching district and building goal(s).	
	• Based on data	
Example:		
	90% of fourth grade students will reach proficiency (80%) on the quarterly assessment given in October 2008.	

D-2

Assessing Your Culture of Quality Data

Why use data? _____

Data Source	Data Tool Used	Content Area	Dates of Collection	Students Assessed	Who uses this data?	How is the data used?	How could we more effectively access and use this data?
MEAP		Reading Writing Total ELA Math	October Results -	Grades 3-8			
MEAP		Science Social Studies	October Results -	Science 5,8 Social Studies 6, 9			
MME		Reading for information ACT plus Writing Work Keys Science Social Studies	March Results -	Grade 11			
District Assessment		Reading					

3d

Data Source	Data Tool Used	Content Area	Dates of Collection	Students Assessed	Who uses this data?	How is the data used?	How could we more effectively access and use this data?
District Assessment		Writing					
District Assessment		Math					
Other							
Other							

*informed from the work of Boudett, City, and Murnane in Data Wise, 2008

What data/information (other than MEAP/MME) does the school use to measure student achievement at each grade level?

What are the criteria for student success at each grade level? What should our students know and be able to do?

What example/exemplars have been developed for analysis of writing and reading?

What processes have been developed for the analysis of writing and reading?

D-5

What instructional initiatives are driven by data?

What do we need to know more about?

CHAPTER 1: DESIGNING A DEPARTMENT ASSESSMENT PLAN

This chapter provides guidelines for designing an assessment plan. It explores the possible players and outlines the important components of an effective plan.

Planning is essential to the success of assessment efforts. It is important to determine what will be done and how the results will be used ***before*** assessment efforts begin. Planning allows for discussion among and involvement of the critical players before any work begins.

Topics Presented

- Definitions of assessment and evaluation
- Key participants in the design of an assessment plan
- Key components of an assessment plan
- Characteristics of an effective assessment plan

Definitions

Assessment is "any effort to gather, analyze, and interpret evidence which describes institutional, divisional, or agency effectiveness" (Upcraft & Schuh, 1996, p. 18).

Evaluation is "any effort to use assessment evidence to improve institutional, departmental, divisional, or institutional effectiveness" (Upcraft & Schuh, 1996, p. 19).

Assessment and evaluation are often combined due to their obvious link. It is important to remember that they are not the same, however. Assessment is the collection of data for the purpose of evaluation. Evaluation is any action taken based upon information from an assessment.

General Information

Participants

Who should participate in designing a department assessment plan?

An assessment ***team*** plans and coordinates assessment efforts. The make-up of the team is important to the ultimate success of the assessment.

Faculty are crucial members of the team. The most effective assessment plan is one that involves and has the support of the faculty. They should be involved in every step of the process from developing assessment objectives to reporting the results. ***Department chairs*** are important because of their role in the implementation of results and the allocation of resources. ***Assessment consultants***, such as the Office of Academic Assessment and Institutional Research, are useful in developing or revising assessment plans, gathering institutional data, completing specialized analyses, assisting with conducting the assessment, or providing professional development support. ***Students, alumni,*** and ***employers*** can serve on committees and provide valuable feedback about assessment activities.

Assessment Plan Components

What should be included in the assessment plan?

An effective assessment plan usually includes the following (Palomba & Banta, 1999, p. 39):

- Subject matter
- Methodology
- Timeline
- Use of assessment information
- Provisions for administering the plan
- Provisions for assessing the assessment

Subject Matter - What are we assessing? Why?

The first step in planning any assessment is to determine its ***purpose***. An assessment can be used to ***document outcomes*** (to determine if a class is meeting general education goals). A department might hold focus groups with its majors to

assess department advising, an assessment activity aimed at ***improvement***. The assessment team must clarify whether the purpose of the assessment is documentation or improvement, because the purpose will guide many of the planning decisions.

Assessment should focus on subject areas that are important to the department (student learning, critical thinking skills, career success of graduates, advising in the program). Any assessment effort is sure to fail if its subject matter is not something that faculty care about. For this reason, assessment is often linked with department goals and objectives. (More information about linking goals and objectives with assessment is included in Chapter 2.)

The assessment team may also need to consider ***other assessment requirements*** such as general education evaluation, Commission for Higher Education review, North Central Association accreditation, professional accreditation, and professional licensing.

Methodology - How will we do the assessment?

This component of an assessment plan should include determining how the assessment will be conducted and identifying the ***assessment tools, data,*** and ***participants***. The methodology should be closely linked with the subject matter. (Chapter 3 discusses selecting assessment tools. Chapters 4 through 8 discuss in depth the specific types of tools, including their advantages and disadvantages.)

Timeline - When will we complete each component?

Specifically, the ***timeline*** will include ***when assessment activities will take place, who will participate,*** and ***who will coordinate the activities***. Assessment activities can be cyclical (annual, bi-annual, etc.) or single instance activities. The timeline is important because it outlines not only when things will occur but who will be involved. Careful attention must be paid to the timeline so that participants and assessment team members are not overloaded.

Using Assessment Information – How will we analyze and use the information?
A number of decisions must be made about how the assessment information will be used, in particular how the data will be analyzed, who will see the results, and what types of reporting will be most effective. The ***type of analysis*** should flow from the purpose of the assessment, the objectives measured, and the tools used. ***Who will see the assessment results*** can be more complicated. Assessment data is often for department use only. Assessment provides an opportunity for the department to take a critical look at its programs, to identify strengths and weaknesses. There are instances when a department is gathering assessment data for a particular audience such as a general education subcommittee, curriculum and advisory committees, accrediting bodies, and the Commission for Higher Education. Finally, what ***type of reporting*** or sharing of information will occur? Short handouts and internal discussions are all that will be needed in some circumstances; official reports are useful or required at other times. (Chapter 9 discusses reporting and using assessment results.)

Provisions for Administering the Plan - Who is responsible for what?
Administration of the plan goes hand in hand with the timeline. The focus is on determining ***who will do what activities.*** Who will instigate the meetings to discuss goals? Who will manage survey distribution? Who will analyze the data? Who will write the final report? Who will share the results?

Assessing the Assessment - How is the plan working?
Ideally, assessment is a process that includes reflection and improvement on assessment efforts. Assessing the plan should include ***evaluating each part of the plan*** from department goals and objectives to use of the information. Although this is often an easy step to overlook, it can be one of the most productive because it guides and improves future assessment efforts.

Frequently Asked Questions

What are the characteristics of an effective assessment plan?

Characteristics of a good assessment program (Palomba & Banta, 1999, p.16)

- Asks important questions
- Reflects institutional mission
- Reflects programmatic goals and objectives for learning
- Contains a thoughtful approach to assessment planning
- Is linked to decision making about the curriculum
- Is linked to processes such as planning and budgeting
- Encourages involvement of individuals from on and off campus
- Contains relevant assessment techniques
- Includes direct evidence of learning
- Reflects what is known about how students learn
- Shares information with multiple audiences
- Leads to reflection and action by faculty, staff, and students
- Allows for continuity, flexibility, and improvement in assessment

Where do we begin?

Start by examining department goals and statements of objectives. What are the intended outcomes of the program? Which of these outcomes are most important to the program? (Chapter 2 discusses goals and objectives and how they can be linked to assessment.)

Topics Reviewed

- Assessment is the gathering, analyzing, and interpreting of data; evaluation is action based upon results of the assessment.
- Faculty, department chairs, assessment consultants, students, alumni, and employers are all possible players in designing a departmental assessment plan.
- The key components of an assessment plan are the subject matter, methodology, timeline, use of assessment information, provisions for administrating the plan, and provisions for assessing the assessment.

Sources of Additional Information

Banta, T. W. (1993). *Making a difference: Outcomes of a decade of assessment in higher education.* San Francisco: Jossey-Bass.

Banta, T. W. (2002). *Building a scholarship of assessment.* San Francisco: Jossey-Bass.

Edwards, A., & Knight, P. (Eds.). (1995). *Assessing competence in higher education.* London: Kogan Page Limited.

Erwin, T. D. (1991). *Assessing student learning and development.* San Francisco: Jossey-Bass.

Freeman, R., & Lewis, R. (1998). *Planning and implementing assessment.* London: Kogan Page Limited.

Nichols, J. O. (1995). *Assessment case studies: Common issues in implementation with various campus approaches to resolution.* Bronx, NY: Agathon Press.

Palomba, C. A., & Banta, T. W. (1999). *Assessment essentials: Planning, implementing, and improving assessment in higher education.* San Francisco: Jossey-Bass.

Stark, J. S., & Thomas, A. (1994). *Assessment program evaluation.* Needham Heights, MA: Simon & Schuster Custom Publishing.

Upcraft, M. L., & Schuh, J. H. (1996). *Assessment in student affairs: A guide for practitioners.* San Francisco: Jossey-Bass.

Assessment Plan (Sample)

Grade	Beginning	Middle	End	Responsible
K: All students	*Comp level-DRA *concepts about print *Letter ID		*Comp level-DRA *Concepts about print	* (Text level used if DRA of 3 or lower)
K: All students			*Letter ID	*
K: All students	Monthly Running Records			*
K: All students	*Ongoing MLPP *Phonological Awareness *Rhyme *Onset and Rime			*
K: All students	*Writing	*Writing	*Writing	*
1: All students	*Comp level-DRA		*Comp level-DRA	*
1:All students	Monthly Running Records			*
1:All students	*Ongoing MLPP *Phoneme Blending *Phoneme Segmentation *Hearing and Recording Sounds *Sight Words (List Provided) *Known Words *Expressive and Receptive Language (Oral Language Section)			*
1: All students	*Letter ID *Concepts about print			*
1:At-risk students			*Concepts about print	*
1: All students	*Writing	*Writing	*Writing	*
2: All students	*Comp level-DRA		*Comp level DRA	*
2: All students	Monthly Running Records			*
2: All students	*Ongoing MLPP *Hearing and Recording Sounds *Sight Words (List Provided) *Known Words *Any Incomplete or not mastered assessments K,1 *Expressive and Receptive Language (Oral language Section)			*
2: All students	*Writing	*Writing	*Writing	*

Grade	Beginning	Middle	End	Responsible
3: All students			*Comp level-DRA	*
3: All students	*Ongoing MLPP *Sight Words (List Provided) *Known Words *Any incomplete or not mastered assessment K,1,2			*
3: All students	*Writing *QRI	*Writing *QRI	*Writing *QRI	*
3: All students	*MEAP – Oct. *MI-Access			*
3: At-risk students (Title) (previous year and current)	*Comp level-DRA			*
4: New students	*Comp level - DRA			*
4: All students	*MEAP – Oct. *MI-Access			*
4: All students	*Writing *QRI	*Writing *QRI	*Writing *QRI	*
4: At-risk students (Title) (previous year and current)	*Comp level-DRA		*Comp level-DRA	*
5: New students	*Comp level-DRA			*
5: At-risk students (Title) (previous year and current)	*Com level-DRA		*Comp level-DRA	*
5: All students	*Writing *QRI	*Writing *QRI	*Writing *QRI	*
5: All students	*MEAP-Oct. *MI-Access			*
6: New students	* Comp level-DRA			*
6: At-risk students (Title) (previous year and current)	*Comp level-DRA		*Comp level-DRA	*
6: All students	*Writing *QRI	*Writing *QRI	*Writing *QRI	*
7: All students	*MEAP-Oct *MI-Access.			*
8: All students	*Writing	*Writing	*Writing	*

	*QRI	*QRI	*QRI	
8: All students	*MEAP-Oct. *MI-Access			*
9: All students	*MEAP-Oct			*
Grade	**Beginning**	**Middle**	**End**	**Responsible**

2008/2009 K-12 Assessment Plan (Sample)

Grade	First Quarter October	Second Quarter January	Third Quarter March	Fourth Quarter June	Responsible
K:All Students					*Teacher
K:All Students					
K:All Students					
K:All Students					
1st:All Students					*Teacher
1st:All Students					
1st:All Students					
1st:All Students					
2nd:All Students					*Teacher
2nd:All Students					
2nd:All Students					
2nd:All Students					
3rd:All Students					*Teacher
3rd	**MEAP** 10/14/08				
3rd	**MI ACCESS** 10/13-11/21/08				
3rd:All Students					
4th:All Students					*Teacher
4th	**MEAP** 10/14/08				
4th	**MI ACCESS** 10/13-11/21/08				
4th:All Students					

2008/2009 K-12 _____ Assessment Plan (Sample)

Grade	First Quarter October	Second Quarter January	Third Quarter March	Fourth Quarter June	Responsible
5th:All Students					*Teacher
5th	**MEAP** 10/14/08				
5th	**MI-ACCESS** 10/13-11/21/08				
5th:All Students					
6th:All Students					*Teacher
6th	**MEAP** 10/14/08				
6th	**MI-ACCESS** 10/13-11/21/08				
6th:All Students					
7th:All Students					*Teacher
7th	**MEAP** 10/14/08				
7th	**MI-ACCESS** 10/13-11/21/08				
7th:All Students					
8th:All Students					*Teacher
8th	**MEAP** 10/14/08				
8th	**MI-ACCESS** 10/13-11/21/08				
8th:All Students					

2008/2009 K-12 Assessment Plan (Sample)

Grade/ Course	First Quarter October	Second Quarter January	Third Quarter March	Fourth Quarter June	Responsible
9th:All Students					*Teacher
9th:All Students					
9th:All Students					
9th:All Students					
10th:All Students					*Teacher
10th:All Students					
10th:All Students					
10th:All Students					
11th:All Students	**MME** 3/12/2009				*Teacher
11th:All Students	**MI-ACCESS** 2/26-3/31/09				
11th:All Students					
11th:All Students					
12th:All Students					*Teacher
12th:All Students					
12th:All Students					
12th:All Students					

Schools Planning Guide

Revised August 8, 2008

Problem Statement

_____ students consistently score below the county and state on the Mathematics MEAP and MME.

- In overall achievement
- In sub-group achievement

Evidence

- See attached data.

_____ District Goal

*Increase overall student achievement in mathematics at each building, as measured on the MEAP or MME by **20%** or reaching 90% proficient level.*

Increase achievement in the following sub-groups by 20% :

- *Special Education students*
- *Black students*
- *Economically Disadvantaged students*

_____ will drive the District Goal.

_____ will support the leadership team in district and building goals.

Buildings will

1. Design a plan to meet the requirements of this goal.
2. Determine SMART goals at grade level to meet this goal.

Have the flexibility to incorporate additional building goals cautiously.

Plan for Reaching Goal

1. Develop a K-12 assessment system for mathematics.
2. Develop pacing guides for each grade level mathematics course.
3. Develop and/or revise quarterly assessments for grades K-8 and as determined appropriate for HS mathematics classes that align to the pacing guides and the GLCE and CCE. Administer according to district mathematics assessment system.
4. Input all assessments in Data warehouse.
5. Design a systematic data analysis process to include data discussions with grade levels/mathematics teachers around the math assessment results planning for student deficits. Focus on sub-groups.
6. Develop a PD plan (from school/district data) to support the mathematics work that needs to be accomplished.
    a. Support teachers in the use/understanding of math curriculum, math concepts, math resources, math instruction, and math vocabulary.

Action Steps	
Strategy Number 1	**Develop a K-12 assessment system for mathematics (Development at the district level; responsibility for implementation will be at the building level)**
Person(s) Responsible	
Persons Involved	
Timeframe Dates for Activities **AND** Dates for Completion	1. **Assessment Calendar - Completed by September 30, 2008.** 2. **Assessment Binder – Completed by March 31, 2009** 3. **Presentation to Board – May 2009**
Resources and Budget	
Evidence of Strategy's Implementation Include data to verify both progress and completion	1. **Completed calendar** 2. **Completed binder** 3. **Presentation to Board** 4. **Buildings implementing assessments according to schedule**
Strategy Number 2	**Develop curriculum maps for each grade level mathematics course (Development at building level w/assistance from district coordinator)**
Person(s) Responsible	
Persons Involved	
Timeframe Dates for Activities **AND** Dates for Completion	1. **Provide teachers with a template that will be used for K-12 curriculum mapping** 2. **Ensure that teachers understand** a. **backwards planning** b. **their GLCE/CCE** c. **that some of the concepts may not be tested at their grade level but they will be expected to begin concept development (vertical alignment)** d. **the intent of curriculum maps** 3. **All curriculum maps (by grade level or course) are completed – September 30, 2008** 4. **Curriculum maps are reviewed for completeness and**

	accuracy – October 15, 2008 5. Staff are presented with and held accountable for the curriculum maps – October 2008
Resources and Budget	
Evidence of Strategy's Implementation Include data to verify both progress and completion	1. Staff understanding of the reason and process for the development of curriculum maps. 2. Curriculum maps completed for all math courses.

Strategy Number 3	**Develop and/or revise quarterly assessments for grades K-8 and as determined appropriate for HS mathematics classes that align to the pacing guides and the GLCE and CCE. (Developed at the building level; reviewed at the district level). Administer according to district mathematics assessment system.**
Person(s) Responsible	
Persons Involved	
Timeframe Dates for Activities **AND** Dates for Completion	1. Ensure that teachers understand a. backwards planning b. their curriculum maps and that the assessments must match the curriculum maps c. their GLCE/CCE and how they determine what is supposed to be assessed. 2. Assessments for quarters one and two completed by October 31st. 3. Assessments reviewed by math curriculum coordinators by October 31 (for first quarter) and November 30 (for second quarter). 4. Assessments for quarters three and four completed by Jan. 14th. 5. Assessments reviewed by math curriculum coordinators by February 15th (for third quarter) and March 15th (for fourth quarter). 6. All assessments ready to be administered by end of each quarter.

	7. **Staff will be trained in accessing assessments from data warehouse. Reminders to teachers of the assessment to be given and when. A plan to hold teachers accountable for following the assessment timeline.**
Resources and Budget	
Evidence of Strategy's Implementation Include data to verify both progress and completion	1. **Assessments for all math courses.** 2. **Quarterly assessment results for every K-12 student taking a math course.**

Strategy Number 4	**Input all assessments in data warehouse.**
Person(s) Responsible	
Persons Involved	
Timeframe Dates for Activities **AND** Dates for Completion	1. **_____ will be responsible for getting the completed assessments to _____ in a timely fashion.** 2. **_____ will enter all assessments into data warehouse so that they are ready to be given at the end of each quarter.** 3. **_____ will help ensure that all staff know how to access assessments in data warehouse.** 4. **_____ will determine data warehouse capabilities for sending out reminders to teachers of when to give the assessment and which one to give.**
Resources and Budget	
Evidence of Strategy's Implementation Include data to verify both progress and completion	1. **All staff will access the quarterly assessments from data warehouse.** 2. **Quarterly assessment results for every K-12 student taking a math course will be available in data warehouse.**

Strategy Number 5	**Design a systematic data analysis process to include data discussions with grade levels/mathematics teachers around the math**

	assessment results planning for student deficits. (Building level)
Person(s) Responsible	
Persons Involved	**All principals**
Timeframe Dates for Activities **AND** Dates for Completion	**1. Develop a district assessment calendar K-12.** **a. Dates** **b. Timeframe – when given, when scanned into data warehouse.** **c. Person responsible** **2. Building level schedule – notification all staff.** **a. Dates** **b. Timeframe – when given, when scanned into data warehouse.** **c. Person responsible** **d. Permissions for viewing data – Building principal, ISD support person, teachers – grade groups?** **3. Schedule data conferences – Building level** **a. Principal /ISD person to review data and plan conference.** **b. Principal and grade level teams meet in data conference.** **c. Plan for intervention developed, evidence determined to be collected to evaluate plan.** **d. Intervention plan implemented and adjusted to meet student learning needs.** **e. Determine PD if necessary.** **4. Schedule data conference with Superintendent** **a. Review building grade level data.** **b. Review intervention plan by grade levels.** **c. Determine PD if necessary.** **Repeat process quarterly**
Resources and Budget	
Evidence of Strategy's Implementation Include data to verify both progress and completion	**Data conference scheduled and held.** **Plan determined and implemented at grade levels.** **Data collected to determine success of intervention.** **Data reviewed on quarterly basis and plan changed to meet**

	student learning needs.
Strategy Number 6	**Develop a PD plan (from school/district data) to support the mathematics work that needs to be accomplished. (Building and district level – coordination)** **Support teachers in the use/understanding of math curriculum, math concepts, math resources, math instruction, and math vocabulary, a systematic data analysis process to include data discussions with grade levels/mathematics teachers around the math assessment results planning for student deficits. (Building level)**
Person(s) Responsible	
Persons Involved	
Timeframe Dates for Activities **AND** Dates for Completion	
Resources and Budget	
Evidence of Strategy's Implementation Include data to verify both progress and completion	

Schools Planning Guide

_____(date)

Problem Statement

-

Evidence

- See attached data.

_____ District Goal

Plan for Reaching Goal

1.

    a.

Action Steps	
Strategy Number 1	
Person(s) Responsible	
Persons Involved	
Timeframe Dates for Activities **AND** Dates for Completion	**1.**
Resources and Budget	
Evidence of Strategy's Implementation Include data to verify both progress and completion	**1.**
Strategy Number 2	
Person(s) Responsible	
Persons Involved	
Timeframe Dates for Activities **AND** Dates for Completion	**1.**
Resources and Budget	
Evidence of Strategy's Implementation Include data to verify both progress and completion	**1.**

Strategy Number 3	
Person(s) Responsible	
Persons Involved	
Timeframe Dates for Activities **AND** Dates for Completion	**1.**
Resources and Budget	
Evidence of Strategy's Implementation Include data to verify both progress and completion	**1.**
Strategy Number 4	
Person(s) Responsible	
Persons Involved	
Timeframe Dates for Activities **AND** Dates for Completion	**1.**
Resources and Budget	
Evidence of Strategy's Implementation Include data to verify both progress and completion	**1.**

Strategy Number 5	
Person(s) Responsible	
Persons Involved	
Timeframe Dates for Activities **AND** Dates for Completion	1.
Resources and Budget	
Evidence of Strategy's Implementation Include data to verify both progress and completion	

Strategy Number 6	
Person(s) Responsible	
Persons Involved	
Timeframe Dates for Activities **AND** Dates for Completion	
Resources and Budget	
Evidence of Strategy's Implementation Include data to verify both progress and completion	

School: _____

Location: _____

Date: _____

Consultant: _____

Notes from School Improvement Plan Review or Last Visit:

Review of data
Progress toward goal

Meeting: August 21, 2008 – Planning meeting for in-service
In-service: August 28, 2008

- Letter of Agreement provided
- MEAP writing data used
- Analyzed writing samples K-5th grade
- Charted strengths of the writer
- Reviewed 2008-2009 Objectives for school year
- Discussed next steps/possible follow-up with staff – vocabulary, how to, GLCE unpacking

D-28

Questions:

Next steps for staff?

10/08

Expectations	Data Used for Planning	Action Taken	Next Steps Timeline	Evidence of Progress Towards Goal	Person Responsible
Administrator's Goal for the Year Teachers understand and use the writing process	MEAP 07/08 writing data used 8/27/08 writing in-service – looking at quality writing	Staff analyzed quality writing at mixed grade level groups Samples used K-5 Learning charted SIP plan sent to ISD Letter of Agreement sent to ISD	Charts typed and copies shared with staff		
Student Achievement Goal for the Year Student growth in writing	SIP – Goal 3 All students will improve skills in paragraph composition		Restate goal as S.M.A.R.T. goal – specific, measurable, achievable, results-oriented, timely (as measured by an assessment)		
Leadership Team/ PLC meeting **Attended by:** • _____ • _____ • • •	Attended planning meeting for writing in-service	8/27 writing in-service planned Copies made of resources	Plan for next steps		

D-29

Expectations	Data Used for Planning	Action Taken	Next Steps Timeline	Evidence of Progress towards Goal	Person Responsible
Leadership Team Action Plan					
Faculty Meeting Agendas Notes					
Plan for Data Collection and Documentation					

D30

Consultant Reflection:

School _____

Location _____

Date _____

Consultant _____

Notes from School Improvement Plan Review or Last Visit:

Review of data
Progress toward goal

Questions:

Expectations	Data for Planning	Action Taken	Next Steps Timeline	Evidence	Person Responsible
Administrator's Goal for the Year					
Student Achievement Goal for the Year					
Leadership Team/ PLC meeting **Attended by:** •					

D-32

Expectations	Data for Planning	Action Taken	Next Steps Timeline	Evidence	Person Responsible
Leadership Team Action Plan					
Faculty Meeting Agendas Notes					
Plan for Data Collection and Documentation					

Consultant Reflection:

Cohort Proficiency - Math

District/Building:

6th Grade 06-07 Year Compared 07/08	Performance Level 4	Performance Level 3	Performance Level 2	Performance Level 1	Totals
Maintained Performance Level	%	%	%	%	Maintained Performance Level %
Declined to this Performance Level	%	%	%		Declined to this Performance Level %
Gained to this Performance Level		%	%	%	Gained to this Performance Level %
					Total %

DJ34

*7th Grade 06-07 Year Compared 07-08	Performance Level 4	Performance Level 3	Performance Level 2	Performance Level 1	Totals
Maintained Performance Level	%	%	%	%	Maintained Performance Level %
Declined to this Performance Level	%	%	%		Declined to this Performance Level %
Gained to this Performance Level		%	%	%	Gained to this Performance Level %
					Total %

D-35

*Cohort Proficiency-Math

11/08

Cohort Proficiency -

District/Building:

Grade Year Compared	Performance Level 4	Performance Level 3	Performance Level 2	Performance Level 1	Totals
Maintained Performance Level	%	%	%	%	Maintained Performance Level %
Declined to this Performance Level	%	%	%	%	Declined to this Performance Level %
Gained to this Performance Level	%	%	%	%	Gained to this Performance Level %
					Total %

4-8

Grade / Year Compared	Performance Level 4	Performance Level 3	Performance Level 2	Performance Level 1	Totals
Maintained Performance Level	%	%	%	%	Maintained Performance Level %
Declined to this Performance Level	%	%	%	%	Declined to this Performance Level %
Gained to this Performance Level	%	%	%	%	Gained to this Performance Level %
					Total %

Q-37

*Cohort Proficiency-

11/08

Cohort Proficiency -

District/Building:

Grade / Year Compared	Performance Level 4	Performance Level 3	Performance Level 2	Performance Level 1	Totals
Maintained Performance Level	%	%	%	%	Maintained Performance Level %
Declined to this Performance Level	%	%	%	%	Declined to this Performance Level %
Gained to this Performance Level	%	%	%	%	Gained to this Performance Level %
					Total %

Grade Year Compared	Performance Level 4	Performance Level 3	Performance Level 2	Performance Level 1	Totals
Maintained Performance Level	%	%	%	%	Maintained Performance Level %
Declined to this Performance Level	%	%	%	%	Declined to this Performance % Level
Gained to this Performance Level	%	%	%	%	Gained to this Performance Level %
					Total %

A-35

*Cohort Proficiency-

11/08

Cohort Proficiency -

District/Building:

Grade Year Compared	Performance Level 4	Performance Level 3	Performance Level 2	Performance Level 1	Totals
Maintained Performance Level	%	%	%	%	Maintained Performance Level %
Declined to this Performance Level	%	%	%	%	Declined to this Performance Level %
Gained to this Performance Level	%	%	%	%	Gained to this Performance Level %
					Total %

Grade Year Compared	Performance Level 4	Performance Level 3	Performance Level 2	Performance Level 1	Totals
Maintained Performance Level	%	%	%	%	Maintained Performance Level %
Declined to this Performance Level	%	%	%	%	Declined to this Performance % Level
Gained to this Performance Level	%	%	%	%	Gained to this Performance Level %
					Total %

*Cohort Proficiency-

11/08

Cohort Proficiency -

District/Building:

Grade / Year Compared	Performance Level 4	Performance Level 3	Performance Level 2	Performance Level 1	Totals
Maintained Performance Level	%	%	%	%	Maintained Performance Level %
Declined to this Performance Level	%	%	%	%	Declined to this Performance Level %
Gained to this Performance Level	%	%	%	%	Gained to this Performance Level %
					Total %

24

Grade Year Compared	Performance Level 4	Performance Level 3	Performance Level 2	Performance Level 1	Totals
Maintained Performance Level	%	%	%	%	Maintained Performance Level %
Declined to this Performance Level	%	%	%	%	Declined to this Performance % Level
Gained to this Performance Level	%	%	%	%	Gained to this Performance Level %
					Total %

DL43

*Cohort Proficiency-

11/08

Analyzing Your MEAP Data: Drilling Down

School _____ Grade _____ Year _____

How did our school perform this year?

Using the MEAP Proficiency Reports, enter the percent of students who were proficient on this year and last year's MEAP.

	Math		ELA		Reading		Writing		Science		Soc Stud	
	06	05	06	05	06	05	06	05	06	05	06	05
School												
District												
ISD												
State												

In what area did we perform the highest?
What was our area of greatest need?
Are we moving in the right direction?
How did our performance compare to district, ISD, and state performance?
Do you see any patterns or trends in this data?

What non-evaluative observations can we make about our data?
What questions do the data raise for us?

D-44

How did our subgroup performance compare?

	English Language Arts				Mathematics			
Group	AYP Target 2006	School % Prof 2006	Gap+ / --	School % Prof 2005	AYP Target 2006	School % prof 2006	Gap + / --	School % Prof 2005
All students								
American Indian								
Asian								
Black								
Hispanic								
White								
Economically disadvantaged								
ELL								
Students with disabilities								

Did we meet our AYP grade level achievement targets?
Are we concerned about any achievement gaps?
Do our state data reflect similar achievement gaps to ours?
Based on this data, what is our greatest need?

What non-evaluative observations can we make about our data?

What questions do the data raise for us?

How did the same students (cohort group) perform this year compared to last year?

Using the Cohort Proficiency Report, identify whether our cohort groups have moved to a higher proficiency level over the past two years.

- Based on this data, are we moving in the right direction?
- If our performance remains the same, what does that mean to you?
- Based on this data, what are your areas of need?
- Click on a bubble on the Cohort Proficiency Report to view the list of students performing at each proficiency level. How would you use this information?

What non-evaluative observations can we make about our data?
What questions do the data raise for us?

How did our students perform on Strands and GLCEs?

Using the Comparative Item Analysis graphs, what can we learn about our instruction from examining student performance on individual items? Print out the graphs that show your performance on each item or use the attached form to complete the following analysis. Highlight on your printout or write on your form the number of any item on which the state outperformed our students by 10 percentage points and the GLCE that was being assessed on that item.

- On which items and GLCEs did the state outperform our students?
- What patterns do you see in the data?
- What additional information can we gain from looking at the questions for these items?

What non-evaluative observations can we make about our data?
What questions do the data raise for us?

School Improvement Planning

Based on your analysis of your MEAP Data, what content area goals do you need to focus on to close the gap between where you are and where you want to be?
Are there specific Strands or GLCEs that you need to monitor next year?
What priority questions did your data raise that you want to explore? What data do you need to help answer those questions?

Approximate Strand Proficiency

To determine an "approximate strand proficiency", we must first create a subset of students whose raw scores will be used to determine a mean score for each strand. This subset is constituted by all the students who are attained a level 2 proficiency across the State for each subject at each grade level. Once we have identified all of the students who meet that criteria, we calculate the mean score for these students at the level of each strand. For our purposes, this is the "approximate strand proficiency".

For your reference, here is the algorithm provided by Joseph Martineau:

- Subset the proficient (level 2) students for each subject by grade.
- Determine the minimum scaled score and raw score for these students (Level 2 for Math Performance Level, etc.)
- Subset the students at the minimum raw score associated with the minimum observed scaled score earned for proficient students.
- Calculate the mean strand score for each strand for this subset of students at each grade for each subject. This mean strand score is the strand score at each grade/subject that will be considered the 'approximate proficiency' score for a strand.
- Create a table that provides the mean strand score for approximate strand proficiency by subject and grade.

Phi Delta Kappan

pdkintl home | site map

Online Article

Assessment Crisis: The Absence Of Assessment *FOR* Learning

If we wish to maximize student achievement in the U.S., we must pay far greater attention to the improvement of classroom assessment, Mr. Stiggins warns. Both assessment of learning and assessment for learning are essential. But one is currently in place, and the other is not.

By Richard J. Stiggins

> A real voyage of discovery consists not of seeking new landscapes but of seeing through new eyes.
>
> -- Marcel Proust

IF WE ARE finally to connect assessment to school improvement in meaningful ways, we must come to see assessment through new eyes. Our failure to find a potent connection has resulted in a deep and intensifying crisis in assessment in American education. Few elected officials are aware of this crisis, and almost no school officials know how to address it. Our current assessment systems are harming huge numbers of students for reasons that few understand. And that harm arises directly from our failure to balance our use of standardized tests and classroom assessments in the service of school improvement. When it comes to assessment, we have been trying to find answers to the wrong questions.

Politicians routinely ask, How can we use assessment as the basis for doling out rewards and punishments to increase teacher and student effort? They want to know how we can intensify the intimidation associated with annual testing so as to force greater achievement. How we answer these questions will certainly affect schools. But that impact will not always be positive. Moreover, politicians who ask such questions typically look past a far more important pair of prior questions: How can we use assessment to help all our students *want* to learn? How can we help them feel *able* to learn? Without answers to these questions, there will be no school improvement. I explain why below.

School administrators in federal, state, and local education agencies contribute to our increasingly damaging assessment crisis when they merely bow to politicians' beliefs and focus unwaveringly on the question of how to make our test scores go up. To be sure, accountability for student learning is important. I am not

D-50

opposed to high-stakes testing to verify school quality -- as long as the tests are of sound quality.1 However, our concern for test scores must be preceded by a consideration of more fundamental questions: Are our current approaches to assessment improving student learning? Might other approaches to assessment have a greater impact? Can we design state and district assessment systems that have the effect of helping our students want to learn and feel able to learn?

Furthermore, the measurement community, of which I am a member, also has missed an essential point. For decades, our priorities have manifested the belief that our job is to discover ever more sophisticated and efficient ways of generating valid and reliable test scores. Again, to be sure, accurate scores are essential. But there remains an unasked prior question: How can we maximize the positive impact of our scores on learners? Put another way, How can we be sure that our assessment instruments, procedures, and scores serve to help learners want to learn and feel able to learn?

We are a nation obsessed with the belief that the path to school improvement is paved with better, more frequent, and more intense standardized testing. The problem is that such tests, ostensibly developed to "leave no student behind," are in fact causing major segments of our student population to be left behind because the tests cause many to give up in hopelessness -- just the opposite effect from that which politicians intended.

Student achievement suffers because these once-a-year tests are incapable of providing teachers with the moment-to-moment and day-to-day information about student achievement that they need to make crucial instructional decisions. Teachers must rely on classroom assessment to do this. The problem is that teachers are unable to gather or effectively use dependable information on student achievement each day because of the drain of resources for excessive standardized testing. There are no resources left to train teachers to create and conduct appropriate classroom assessments. For the same reasons, district and building administrators have not been trained to build assessment systems that balance standardized tests and classroom assessments. As a direct result of these chronic, long-standing problems, our classroom, building, district, state, and national assessment systems remain in constant crisis, and students suffer the consequences. All school practitioners know this, yet almost no politicians do.

We know how to build healthy assessment environments that can meet the information needs of all instructional decision makers, help students want to learn and feel able to learn, and thus support unprecedented increases in student achievement. But to achieve this goal, we must put in place the mechanisms that will make healthy assessment possible. Creating those mechanisms will require that we begin to see assessment through new eyes. The well-being of our students depends on our willingness to do so.

The Evolution of Our Vision of Excellence in Assessment

The evolution of assessment in the United States over the past five decades has led to the strongly held view that school improvement requires:

* the articulation of higher achievement *standards*,
* the transformation of those expectations into rigorous *assessments*, and
* the expectation of *accountability* on the part of educators for student achievement, as reflected in test scores.

Standards frame accepted or valued definitions of academic success. Accountability compels attention to these standards as educators plan and deliver instruction in the classroom. Assessment provides the evidence of success on the part of students, teachers, and the system.

To maximize the energy devoted to school improvement, we have "raised the bar" by setting world-class standards for student achievement, as opposed to minimum competencies. To further intensify the impact of our standards and assessments, policy makers often attach the promise of rewards for schools that produce high scores and sanctions for schools that do not.

In this context, we rely on high-stakes assessments *of learning* to inform our decisions about accountability. These tests tell us how much students have learned, whether standards are being met, and whether educators have done the job they were hired to do.

Such assessments of learning have been the norm throughout the U.S. for decades. We began with standardized college admissions tests in the early decades of the last century, and this use of testing continues essentially unchanged today. But these tests are not used merely for college admission. For decades, we have ranked states according to average SAT scores.

Meanwhile, in response to demands for accountability in public schools in the 1960s, we launched districtwide standardized testing programs that also remain in place today. In the 1970s, we began the broad implementation of statewide testing programs, and these programs have spread throughout the land. Also in the 1970s and extending into the 1980s, we added a national assessment program that continues to this day. During the 1990s, we became deeply involved and invested in international assessment programs. Across the nation, across the various levels of schooling, and over the decades, we have invested billions of dollars to ensure the accuracy of the scores on these assessments of learning. Now in 2002, President Bush has signed a school reform measure that requires standardized testing of every pupil in the U.S. in mathematics and reading every year in grades 3 through 8, once again revealing our faith in assessment as a tool for school improvement.

In the context of school improvement, we have seen assessment merely as an index of the success of our efforts. It is testimony to our societal belief in the power of standardized tests that we would permit so many levels of testing to remain in place, all at the same time and at very high cost. Clearly, over the decades, we have believed that by checking achievement status and reporting the results to the public we can apply the pressure needed to intensify -- and thus speed -- school improvement. At the same time, we have believed that providing policy makers and practicing educators with test results can inform the critically important school improvement decisions that are made at district, state, and federal levels.

The Flaw in the Vision

The assessment environment described above is a direct manifestation of a set of societal beliefs about what role assessment ought to play in American schools. Over the decades, we have succeeded in carrying these beliefs to unfortunate extremes.

For example, we have believed that assessment should serve two purposes: inform decisions and motivate learning. With respect to the former, we have built our assessment systems around the belief that the most important decisions are made by those program planners and policy makers whose actions affect the broadest range of classrooms and students. The broader the reach of the decision makers (across an entire school district or state), the more weight we have given to meeting their information needs first. This is the foundation of our strong belief in the power of standardized tests. These are the tests that provide comparable

data that can be aggregated across schools, districts, and states to inform far-reaching programmatic decisions.

With respect to the use of assessment to motivate, we all grew up in classrooms in which our teachers believed that the way to maximize learning was to maximize anxiety, and assessment has always been the great intimidator. Because of their own very successful experiences in ascending to positions of leadership and authority, most policy makers and school leaders share the world view that, "when the going gets tough, the tough get going." They learned that the way to succeed when confronted with a tougher challenge is to redouble your efforts -- work harder and work smarter. If you do so, you win. And so, they contend, the way to cause students to learn more -- and thus the way to improve schools -- is to confront them with a tougher challenge. This will cause them to redouble their efforts, they will learn more, their test scores will go up, and the schools will become more effective. We can motivate students to greater effort, they believe, by "setting higher academic standards," "raising the bar," and implementing more high-stakes testing. This is the foundation of our belief in the power of accountability-oriented standardized tests to drive school improvement.

In point of fact, when some students are confronted with the tougher challenge of high-stakes testing, they do redouble theirefforts, and they do learn more than they would have without the added incentive. Please note, however, that I said this is true for "some students."

Another huge segment of our student population, when confronted with an even tougher challenge than the one that it has already been failing at, will not redouble its efforts -- a point that most people are missing. These students will see both the new high standards and the demand for higher test scores as unattainable for them, and they will *give up in hopelessness*.

Many political and school leaders have never experienced the painful, embarrassing, and discouraging trauma of chronic and public academic failure. As a result, they have no way of anticipating or understanding how their high-stakes testing program, whether local or statewide, could lead to even greater failure for large numbers of students. But tapping the intimidation power of standardized tests for public accountability has an effect on the success of this segment of the student population that is exactly the opposite of what we intend.

Thus it is folly to build our assessment environments on the assumption that standardized testing will have the same effect on all students. It will not. Some students approach the tests with a strong personal academic history and an expectation of success. Others approach them with a personal history and expectation of very painful failure. Some come to slay the dragon, while others expect to be devoured by it. As a result, high-stakes assessment will enhance the learning of some while discouraging others and causing them to give up. Yet, as they attempt to weave assessment into the school improvement equation, federal, state, and local policy makers seem unable to understand or to accommodate this difference.

A More Powerful Vision

There is another way in which assessment can contribute to the development of effective schools that has been largely ignored in the evolution of the standards, assessment, and accountability movement described above. We can also use assessments *for learning*.2 If assessments *of learning* provide evidence of achievement for public reporting, then assessments *for learning* serve to help students learn more. The crucial distinction is between assessment to determine the status of learning and assessment to promote greater learning.

Assessments *of* and *for* learning are both important. Since we in the U.S. already have many assessments *of* learning in place, if we are to balance the two, we must make a much stronger investment in assessment *for* learning. We can realize unprecedented gains in achievement if we turn the current day-to-day classroom assessment process into a more powerful tool for learning. We know that schools will be held accountable for raising test scores. Now we must provide teachers with the assessment tools needed to do the job.

It is tempting to equate the idea of assessment *for learning* with our more common term, "formative assessment." But they are not the same. Assessment *for learning* is about far more than testing more frequently or providing teachers with evidence so that they can revise instruction, although these steps are part of it. In addition, we now understand that assessment *for learning* must involve students in the process.

When they assess *for learning*, teachers use the classroom assessment process and the continuous flow of information aboutstudent achievement that it provides in order to advance, not merely check on, student learning. They do this by:

* understanding and articulating *in advance of teaching* the achievement targets that their students are to hit;
* informing their students about those learning goals, *in terms that students understand*, from the very beginning of the teaching and learning process;
* becoming assessment literate and thus able to transform their expectations into assessment exercises and scoring procedures that *accurately reflect student achievement*;
* using classroom assessments to *build students' confidence* in themselves as learners and help them take responsibility for their own learning, so as to lay a foundation for lifelong learning;
* translating classroom assessment results into frequent *descriptive feedback* (versus judgmental feedback) for students, providing them with specific insights as to how to improve;
* continuously *adjusting instruction* based on the results of classroom assessments;
* engaging students in *regular self-assessment*, with standards held constant so that students can watch themselves grow over time and thus feel in charge of their own success; and
* actively involving students in *communicating* with their teacher and their families about their achievement status and improvement.

In short, the effect of assessment *for learning*, as it plays out in the classroom, is that students keep learning and remain confident that they can continue to learn at productive levels if they keep trying to learn. In other words, students don't give up in frustration or hopelessness.

Are Teachers Ready?

Few teachers are prepared to face the challenges of classroom assessment because they have not been given the opportunity to learn to do so. It is currently the case that only about a dozen states explicitly require competence in assessment as a condition to be licensed to teach. Moreover, there is no licensing examination in place at the state or federal level in the U.S. that verifies competence in assessment. Thus teacher preparation programs have taken little note of competence in assessment, and the vast majority of programs fail to provide the assessment literacy required to enable teachers to engage in assessment *for learning*. It has been so for decades.

Furthermore, lest we believe that teachers can turn to their principals for help, it is currently the case that almost no states require competence in assessment as a condition to be licensed as a principal or school

administrator at any level. Consequently, assessment training is almost nonexistent in administrator training programs. It has been so for decades.

Thus we remain a national faculty that is unschooled in the principles of sound assessment -- whether assessment *of* or *for* learning. This fact has been a matter of record for decades. To date, as a nation, we have invested almost nothing in assessment *for learning*. Teachers rarely have the opportunity to learn how to use assessment as a teaching and learning tool. And our vigorous efforts to assess learning through our various layers of standardized tests cannot overcome the effects of this reality.

As a result of this state of affairs, we face the danger that student progress may be mismeasured, day to day, in classrooms across the nation. That means that all the critically important day-to-day instructional decisions made by students, teachers, and parents may be based on misinformation about student success. The result is the misdiagnosis of student needs, students' misunderstanding of their own ability to learn, miscommunication to parents and others about student progress, and virtually no effective assessment *for learning* in classrooms. The extremely harmful consequences for student learning are obvious.

Relevant Position Statements

The dire consequences of this assessment crisis and the urgent need for action have not gone unnoticed. For example, during the 1990s, virtually every professional association that had anything to do with teaching adopted standards of professional competence for teachers that include an assessment component.3 This group included the American Federation of Teachers (AFT), the National Education Association (NEA), the Council of Chief State School Officers, the National Board for Professional Teaching Standards, and the National Council on Measurement in Education (NCME).

The documents that were issued included a collaborative statement of assessment competencies for teachers developed by a joint committee representing AFT, NEA, and NCME.4 In addition to other standards, this joint statement expects teachers to be trained to choose and develop proper assessment methods; to administer, score, and interpret assessment results; to connect those results to specific decisions; to assign grades appropriately; and to communicate effectively about student achievement. It is troubling to realize that these standards are more than a decade old and still have had little impact on the preparation of teachers and administrators.

In its 2001 report, the Committee on the Foundations of Assessment of the National Research Council advanced recommendations for the development of assessment in American schools that included the following:

Recommendation 9: Instruction in how students learn and how learning can be assessed should be a major component of teacher preservice and professional development programs. This training should be linked to actual experience in classrooms in assessing and interpreting the development of student competence. To ensure that this occurs, state and national standards for teacher licensure and program accreditation should include specific requirements focused on the proper integration of learning and assessment in teachers' educational experience.5

* * *

*Recommendation 11: The balance of mandates and resources should be shifted from an emphasis on external forms of assessment to an increased emphasis on classroom formative assessment designed to assist learning.*6

Similarly, the Commission on Instructionally Supportive Assessment convened by the American Association of School Administrators, the National Association of Elementary School Principals, the National Association of Secondary School Principals, the NEA, and the National Middle School Association included the following in its list of nine requirements for state-mandated accountability tests:

*Requirement 8: A state must ensure that educators receive professional development focused on how to optimize children's learning based on the results of instructionally supportive assessment.*7

We understand what teachers need to know and the proficiencies that they need to develop in order to be able to establish and maintain productive assessment environments. The challenge we face is to provide the opportunity for teachers to master those essential classroom assessment competencies. The depth of this challenge becomes clear when we realize that we must provide opportunities both for new teachers to gain these competencies before they enter the classroom and for experienced teachers who had no chance to master them during their training to gain them as well.

Balancing Assessments *of* and *for* Learning

Therefore, our national assessment priority should be to make certain that assessments both *of* and *for* learning are accurate in their depiction of student achievement and are used to benefit students. Since our standardized assessments *of learning* have been developed by professionals and are currently in place, they are poised to detect any improvements in the level or rate of student achievement.

But these tests provide information only once a year, and we must not delude ourselves into believing that they can serve allassessment purposes. They can reflect large-group increases or decreases in learning on an annual basis, and they can serve as gatekeepers for high-stakes decisions. They cannot inform the moment-to-moment, day-to-day, and week-to-week instructional decisions faced by students and teachers seeking to manage the learning process as it unfolds. They cannot diagnose student needs during learning, tell students what study tactics are or are not working, or keep parents informed about how to support the work of their children. These kinds of uses require assessments *for learning*. The critical question for school improvement is, What would happen to standardized test scores if we brought assessments *for learning* online as a full partner in support of student learning? Several published reviews of research reveal the startling and very encouraging answer.

In 1984 Benjamin Bloom provided a summary of research comparing standard whole-class instruction (the control condition) with two experimental interventions, a mastery learning environment and one-on-one tutoring of individual students. One hallmark of both experimental conditions was the extensive use of classroom assessment *for learning* as a key part of the instructional process. The analyses revealed

differences ranging from one to two standard deviations in student achievement attributable to differences between experimental and control conditions.8

In their 1998 research review, Paul Black and Dylan Wiliam examined the research literature on assessment worldwide, asking if improved formative (i.e., classroom) assessments yield higher student achievement as reflected in summative assessments. If so, they asked, what kinds of improvements in classroom assessment practice are likely to yield the greatest gains in achievement?

Black and Wiliam uncovered and then synthesized more than 250 articles that addressed these issues. Of these, several dozen directly addressed the question of the impact on student learning with sufficient scientific rigor and experimental control to permit firm conclusions. Upon pooling the information on the estimated effects of improved formative assessment on summative test scores, they reported unprecedented positive effects on student achievement. They reported effect sizes of one-half to a full standard deviation. Furthermore, Black and Wiliam reported that "improved formative assessment helps low achievers more than other students and so reduces the range of achievement while raising achievement overall."9 This result has direct implications for districts seeking to reduce achievement gaps between minorities and other students. Hypothetically, if assessment *for learning*, as described above, became standard practice only in classrooms of low-achieving, low-socioeconomic-status students, the achievement gaps that trouble us so deeply today would be erased. I know of no other school improvement innovation that can claim effects of this nature or size.

To fully appreciate the magnitude of the effect sizes cited above, readers need to understand that a gain of one standard deviation, applied to the middle of the test score distribution on commonly used standardized achievement tests, can yield average gains of more than 30 percentile points, two grade-equivalents, or 100 points on the SAT scale. Black and Wiliam report that gains of this magnitude, if applied to the most recent results of the Third International Mathematics and Science Study, would have raised a nation in the middle of the pack among the 42 participating countries (where the U.S. is ranked) to the top five.

This research reveals that these achievement gains are maximized in contexts where educators increase the accuracy of classroom assessments, provide students with frequent informative feedback (versus infrequent judgmental feedback), and involve students deeply in the classroom assessment, record keeping, and communication processes. In short, these gains are maximized where teachers apply the principles of assessment *for learning*.

Black and Wiliam conclude their summary of self-assessment by students as follows:

Thus self-assessment by pupils, far from being a luxury, is in fact *an essential component of formative assessment*. When anyone is trying to learn, feedback about the effort has three elements: redefinition of the *desired goal*, evidence about *present position*, and some understanding of a *way to close the gap between the two*. All three must be understood to some degree by anyone before he or she can take action to improve learning.10 (Emphasis in original.)

Anticipating the Benefits of Balance

Students benefit from assessment *for learning* in several critical ways. First, they become more confident learners because they get to watch themselves succeeding. This success permits them to take the risk of

continuing to try to learn. The result is greater achievement for all students -- especially low achievers, which helps reduce the achievement gap between middle-class and low-socioeconomic-status students. Furthermore, students come to understand what it means to be in charge of their own learning -- to monitor their own success and make decisions that bring greater success. This is the foundation of lifelong learning.

Teachers benefit because their students become more motivated to learn. Furthermore, their instructional decisions are informed by more accurate information about student achievement. Teachers also benefit from the savings in time that result from their ability to develop and use classroom assessments more efficiently.

Parents benefit as well in seeing higher achievement and greater enthusiasm for learning in their children. They also come to understand that their children are learning to manage their own lifelong learning.

School administrators and instructional leaders benefit from the reality of meeting accountability standards and from the public recognition of doing so. Political officials benefit in the same way. When schools work more effectively, both political leaders and school leaders are recognized as contributing to that outcome.

In short, everyone wins. There are no losers. But the price that we must pay to achieve such benefits is an investment in teachers and their classroom assessment practices. We must initiate a program of professional development specifically designed to give teachers the expertise they need to assess *for learning*.

An Action Plan

If we wish to maximize student achievement in the U.S., we must pay far greater attention to the improvement of classroomassessment. Both assessment *of learning* and assessment *for learning* are essential. One is in place; the other is not. Therefore, we must:

* match every dollar invested in instruments and procedures intended for assessment *of learning* at national, state, and local levels with another dollar devoted to the development of assessment *for learning*;
* launch a comprehensive, long-term professional development program at the national, state, and local levels to foster literacy in classroom assessment for teachers, allocating sufficient resources to provide them with the opportunity to learn and grow professionally;
* launch a similar professional development program in effective large-scale and classroom assessment for state, district, and building administrators, teaching them how to provide leadership in this area of professional practice;
* change teacher and administrator licensing standards in every state and in all national certification contexts to reflect an expectation of competence in assessment both *of* and *for* learning; and
* require all teacher and administrator preparation programs to ensure that graduates are assessment literate -- in terms both of promoting and of documenting student learning.

Federal education officials, state policy makers, and local school leaders must allocate resources in equal proportions to ensure the accuracy and effective use of assessments both *of* and *for* learning. Only then can we reassure families that their children are free from the harm that results from the mismeasurement of their achievement in schools. Only then can we maximize students' confidence in themselves as learners. Only then can we raise achievement levels for all students and "leave no child behind."

---

1. For specific standards of quality, refer to Commission on Instructionally Supportive Assessment, *Building Tests to Support Instruction and Accountability* (Washington, D.C.: AASA, NAESP, NASSP, NEA, and NMSA, 2001).

2. This term was coined by Assessment Reform Group, *Assessment for Learning: Beyond the Black Box* (Cambridge: School of Education, Cambridge University, 1999).

3. See the special section on Quality Teaching for the 21st Century in the November 1996 *Phi Delta Kappan*, pp. 190-227.

4. American Federation of Teachers, National Council on Measurement in Education, and National Education Association, "Standards for Teacher Competence in Educational Assessment of Students," *Educational Measurement: Issues and Practice*, vol. 9, no. 4, 1990, pp. 30-32.

5. James W. Pellegrino, Naomi Chudowsky, and Robert Glaser, eds., *Knowing What Students Know: The Science and Design of Educational Assessment* (Washington, D.C.: National Academy Press, 2001), p. 14.

6. Ibid.

7. Commission on Instructionally Supportive Assessment, p. 25.

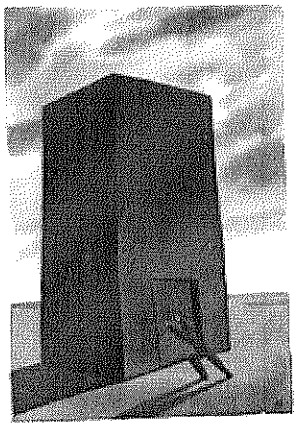
8. Benjamin Bloom, "The Search for Methods of Group Instruction as Effective as One-on-One Tutoring," *Educational Leadership*, May 1984, pp. 4-17.

9. Paul Black and Dylan Wiliam, "Inside the Black Box: Raising Standards Through Classroom Assessment," *Phi Delta Kappan*, October 1998, p. 141. Their work is reported in more detail in idem, "Assessment and Classroom Learning," *Assessment in Education*, March 1998, pp. 7-74.

10. Black and Wiliam, "Inside the Black Box," p. 143.

RICHARD J. STIGGINS is president of Assessment Training Institute, Inc., Portland, Ore.

Kappan Professional Journal
Last updated 6 June 2002
URL: http://www.pdkintl.org/kappan/k0206sti.htm



ONLINE ARTICLE

Inside the Black Box: Raising Standards Through Classroom Assessment

By Paul Black and Dylan Wiliam

Firm evidence shows that formative assessment is an essential component of classroom work and that its development can raise standards of achievement, Mr. Black and Mr. Wiliam point out. Indeed, they know of no other way of raising standards for which such a strong prima facie case can be made.

RAISING the standards of learning that are achieved through schooling is an important national priority. In recent years, governments throughout the world have been more and more vigorous in making changes in pursuit of this aim. National, state, and district standards; target setting; enhanced programs for the external testing of students' performance; surveys such as NAEP (National Assessment of Educational Progress) and TIMSS (Third International Mathematics and Science Study); initiatives to improve school planning and management; and more frequent and thorough inspection are all means toward the same end. But the sum of all these reforms has not added up to an effective policy because something is missing.

Learning is driven by what teachers and pupils do in classrooms. Teachers have to manage complicated and demanding situations, channeling the personal, emotional, and social pressures of a group of 30 or more youngsters in order to help them learn immediately and become better learners in the future. Standards can be raised only if teachers can tackle this task more effectively. What is missing from the efforts alluded to above is any direct help with this task. This fact was recognized in the TIMSS video study: "A focus on standards and accountability that ignores the processes of teaching and learning in classrooms will not provide the direction that teachers need in their quest to improve."[1]

In terms of systems engineering, present policies in the U.S. and in many other countries seem to treat the classroom as a black box. Certain *inputs* from the outside -- pupils, teachers, other resources, management rules and requirements, parental anxieties, standards, tests with high stakes, and so on -- are fed into the box. Some *outputs* are

supposed to follow: pupils who are more knowledgeable and competent, better test results, teachers who are reasonably satisfied, and so on. But what is happening inside the box? How can anyone be sure that a particular set of new inputs will produce better outputs if we don't at least study what happens inside? And why is it that most of the reform initiatives mentioned in the first paragraph are not aimed at giving direct help and support to the work of teachers in classrooms?

The answer usually given is that it is up to teachers: they have to make the inside work better. This answer is not good enough, for two reasons. First, it is at least possible that some changes in the inputs may be counterproductive and make it harder for teachers to raise standards. Second, it seems strange, even unfair, to leave the most difficult piece of the standards-raising puzzle entirely to teachers. If there are ways in which policy makers and others can give direct help and support to the everyday classroom task of achieving better learning, then surely these ways ought to be pursued vigorously.

This article is about the inside of the black box. We focus on one aspect of teaching: formative assessment. But we will show that this feature is at the heart of effective teaching.

The Argument

We start from the self-evident proposition that teaching and learning must be interactive. Teachers need to know about their pupils' progress and difficulties with learning so that they can adapt their own work to meet pupils' needs -- needs that are often unpredictable and that vary from one pupil to another. Teachers can find out what they need to know in a variety of ways, including observation and discussion in the classroom and the reading of pupils' written work.

We use the general term *assessment* to refer to all those activities undertaken by teachers -- and by their students in assessing themselves -- that provide information to be used as feedback to modify teaching and learning activities. Such assessment becomes *formative assessment* when the evidence is actually used to adapt the teaching to meet student needs.[2]

There is nothing new about any of this. All teachers make assessments in every class they teach. But there are three important questions about this process that we seek to answer:

- Is there evidence that improving formative assessment raises standards?
- Is there evidence that there is room for improvement?
- Is there evidence about how to improve formative assessment?

In setting out to answer these questions, we have conducted an extensive survey of the research literature. We have checked through many books and through the past nine years' worth of issues of more than 160 journals, and we have studied earlier reviews of research. This process yielded about 580 articles or chapters to study. We prepared a lengthy review, using material from 250 of these sources, that has been published in a

special issue of the journal *Assessment in Education*, together with comments on our work by leading educational experts from Australia, Switzerland, Hong Kong, Lesotho, and the U.S.3

The conclusion we have reached from our research review is that the answer to each of the three questions above is clearly yes. In the three main sections below, we outline the nature and force of the evidence that justifies this conclusion. However, because we are presenting a summary here, our text will appear strong on assertions and weak on the details of their justification. We maintain that these assertions are backed by evidence and that this backing is set out in full detail in the lengthy review on which this article is founded.

We believe that the three sections below establish a strong case that *governments, their agencies, school authorities, and the teaching profession should study very carefully whether they are seriously interested in raising standards in education*. However, we also acknowledge widespread evidence that fundamental change in education can be achieved only slowly -- through programs of professional development that build on existing good practice. Thus we do not conclude that formative assessment is yet another "magic bullet" for education. The issues involved are too complex and too closely linked to both the difficulties of classroom practice and the beliefs that drive public policy. In a final section, we confront this complexity and try to sketch out a strategy for acting on our evidence.

Does Improving Formative Assessment Raise Standards?

A research review published in 1986, concentrating primarily on classroom assessment work for children with mild handicaps, surveyed a large number of innovations, from which 23 were selected.4 Those chosen satisfied the condition that quantitative evidence of learning gains was obtained, both for those involved in the innovation and for a similar group not so involved. Since then, many more papers have been published describing similarly careful quantitative experiments. Our own review has selected at least 20 more studies. (The number depends on how rigorous a set of selection criteria are applied.) All these studies show that innovations that include strengthening the practice of formative assessment produce significant and often substantial learning gains. These studies range over age groups from 5-year-olds to university undergraduates, across several school subjects, and over several countries.

For research purposes, learning gains of this type are measured by comparing the average improvements in the test scores of pupils involved in an innovation with the range of scores that are found for typical groups of pupils on these same tests. The ratio of the former divided by the latter is known as the *effect size*. Typical effect sizes of the formative assessment experiments were between 0.4 and 0.7. These effect sizes are larger than most of those found for educational interventions. The following examples illustrate some practical consequences of such large gains.

- An effect size of 0.4 would mean that the average pupil involved in an innovation would record the same achievement as a pupil in the top 35% of those not so involved.
- An effect size gain of 0.7 in the recent international comparative studies in mathematics5 would have raised the score of a nation in the middle of the pack of 41 countries (e.g., the U.S.) to one of the top five.

Many of these studies arrive at another important conclusion: that improved formative assessment helps low achievers more than other students and so reduces the range of achievement while raising achievement overall. A notable recent example is a study devoted entirely to low-achieving students and students with learning disabilities, which shows that frequent assessment feedback helps both groups enhance their learning.6 Any gains for such pupils could be particularly important. Furthermore, pupils who come to see themselves as unable to learn usually cease to take school seriously. Many become disruptive; others resort to truancy. Such young people are likely to be alienated from society and to become the sources and the victims of serious social problems.

Thus it seems clear that very significant learning gains lie within our grasp. The fact that such gains have been achieved by a variety of methods that have, as a common feature, enhanced formative assessment suggests that this feature accounts, at least in part, for the successes. However, it does not follow that it would be an easy matter to achieve such gains on a wide scale in normal classrooms. Many of the reports we have studied raise a number of other issues.

- All such work involves new ways to enhance feedback between those taught and the teacher, ways that will require significant changes in classroom practice.
- Underlying the various approaches are assumptions about what makes for effective learning -- in particular the assumption that students have to be actively involved.
- For assessment to function formatively, the results have to be used to adjust teaching and learning; thus a significant aspect of any program will be the ways in which teachers make these adjustments.
- The ways in which assessment can affect the motivation and self-esteem of pupils and the benefits of engaging pupils in self-assessment deserve careful attention.

Is There Room for Improvement?

A poverty of practice. There is a wealth of research evidence that the everyday practice of assessment in classrooms is beset with problems and shortcomings, as the following selected quotations indicate.

- "Marking is usually conscientious but often fails to offer guidance on how work can be improved. In a significant minority of cases, marking reinforces underachievement and underexpectation by being too generous or unfocused. Information about pupil performance received by the teacher is insufficiently used

to inform subsequent work," according to a United Kingdom inspection report on secondary schools.7

- "Why is the extent and nature of formative assessment in science so impoverished?" asked a research study on secondary science teachers in the United Kingdom.8
- "Indeed they pay lip service to [formative assessment] but consider that its practice is unrealistic in the present educational context," reported a study of Canadian secondary teachers.9
- "The assessment practices outlined above are not common, even though these kinds of approaches are now widely promoted in the professional literature," according to a review of assessment practices in U.S. schools.10

The most important difficulties with assessment revolve around three issues. The first issue is *effective learning.*

- The tests used by teachers encourage rote and superficial learning even when teachers say they want to develop understanding; many teachers seem unaware of the inconsistency.
- The questions and other methods teachers use are not shared with other teachers in the same school, and they are not critically reviewed in relation to what they actually assess.
- For primary teachers particularly, there is a tendency to emphasize quantity and presentation of work and to neglect its quality in relation to learning.

The second issue is *negative impact.*

- The giving of marks and the grading function are overemphasized, while the giving of useful advice and the learning function are underemphasized.
- Approaches are used in which pupils are compared with one another, the prime purpose of which seems to them to be competition rather than personal improvement; in consequence, assessment feedback teaches low-achieving pupils that they lack "ability," causing them to come to believe that they are not able to learn.

The third issue is the *managerial role* of assessments.

- Teachers' feedback to pupils seems to serve social and managerial functions, often at the expense of the learning function.
- Teachers are often able to predict pupils' results on external tests because their own tests imitate them, but at the same time teachers know too little about their pupils' learning needs.
- The collection of marks to fill in records is given higher priority than the analysis of pupils' work to discern learning needs; furthermore, some teachers pay no attention to the assessment records of their pupils' previous teachers.

Of course, not all these descriptions apply to all classrooms. Indeed, there are many schools and classrooms to which they do not apply at all. Nevertheless, these general conclusions have been drawn by researchers who have collected evidence -- through observation, interviews, and questionnaires -- from schools in several countries, including the U.S.

An empty commitment. The development of national assessment policy in England and Wales over the last decade illustrates the obstacles that stand in the way of developing policy support for formative assessment. The recommendations of a government task force in 198811 and all subsequent statements of government policy have emphasized the importance of formative assessment by teachers. However, the body charged with carrying out government policy on assessment had no strategy either to study or to develop the formative assessment of teachers and did no more than devote a tiny fraction of its resources to such work.12 Most of the available resources and most of the public and political attention were focused on national external tests. While teachers' contributions to these "summative assessments" have been given some formal status, hardly any attention has been paid to their contributions through formative assessment. Moreover, the problems of the relationship between teachers' formative and summative roles have received no attention.

It is possible that many of the commitments were stated in the belief that formative assessment was not problematic, that it already happened all the time and needed no more than formal acknowledgment of its existence. However, it is also clear that the political commitment to external testing in order to promote competition had a central priority, while the commitment to formative assessment was marginal. As researchers the world over have found, high-stakes external tests always dominate teaching and assessment. However, they give teachers poor models for formative assessment because of their limited function of providing overall summaries of achievement rather than helpful diagnosis. Given this fact, it is hardly surprising that numerous research studies of the implementation of the education reforms in the United Kingdom have found that formative assessment is "seriously in need of development."13 With hindsight, we can see that the failure to perceive the need for substantial support for formative assessment and to take responsibility for developing such support was a serious error.

In the U.S. similar pressures have been felt from political movements characterized by a distrust of teachers and a belief that external testing will, on its own, improve learning. Such fractured relationships between policy makers and the teaching profession are not inevitable -- indeed, many countries with enviable educational achievements seem to manage well with policies that show greater respect and support for teachers. While the situation in the U.S. is far more diverse than that in England and Wales, the effects of high-stakes state-mandated testing are very similar to those of the external tests in the United Kingdom. Moreover, the traditional reliance on multiple-choice testing in the U.S. -- not shared in the United Kingdom -- has exacerbated the negative effects of such policies on the quality of classroom learning.

How Can We Improve Formative Assessment?

The self-esteem of pupils. A report of schools in Switzerland states that "a number of pupils . . . are content to 'get by.' . . . Every teacher who wants to practice formative assessment must reconstruct the teaching contracts so as to counteract the habits acquired by his pupils."14

The ultimate user of assessment information that is elicited in order to improve learning is the pupil. There are negative and positive aspects of this fact. The negative aspect is illustrated by the preceding quotation. When the classroom culture focuses on rewards, "gold stars," grades, or class ranking, then pupils look for ways to obtain the best marks rather than to improve their learning. One reported consequence is that, when they have any choice, pupils avoid difficult tasks. They also spend time and energy looking for clues to the "right answer." Indeed, many become reluctant to ask questions out of a fear of failure. Pupils who encounter difficulties are led to believe that they lack ability, and this belief leads them to attribute their difficulties to a defect in themselves about which they cannot do a great deal. Thus they avoid investing effort in learning that can lead only to disappointment, and they try to build up their self-esteem in other ways.

The positive aspect of students' being the primary users of the information gleaned from formative assessments is that negative outcomes -- such as an obsessive focus on competition and the attendant fear of failure on the part of low achievers -- are not inevitable. What is needed is a culture of success, backed by a belief that all pupils can achieve. In this regard, formative assessment can be a powerful weapon if it is communicated in the right way. While formative assessment can help all pupils, it yields particularly good results with low achievers by concentrating on specific problems with their work and giving them a clear understanding of what is wrong and how to put it right. Pupils can accept and work with such messages, provided that they are not clouded by overtones about ability, competition, and comparison with others. In summary, the message can be stated as follows: *feedback to any pupil should be about the particular qualities of his or her work, with advice on what he or she can do to improve, and should avoid comparisons with other pupils.*

Self-assessment by pupils. Many successful innovations have developed self- and peer-assessment by pupils as ways of enhancing formative assessment, and such work has achieved some success with pupils from age 5 upward. This link of formative assessment to self-assessment is not an accident; indeed, it is inevitable.

To explain this last statement, we should first note that the main problem that those who are developing self-assessments encounter is not a problem of reliability and trustworthiness. Pupils are generally honest and reliable in assessing both themselves and one another; they can even be too hard on themselves. The main problem is that pupils can assess themselves only when they have a sufficiently clear picture of the targets that their learning is meant to attain. Surprisingly, and sadly, many pupils do not have such a picture, and they appear to have become accustomed to receiving classroom teaching as an arbitrary sequence of exercises with no overarching rationale. To overcome this pattern of passive reception requires hard and sustained work. When pupils do acquire such an overview, they then become more committed and more effective as learners.

Moreover, their own assessments become an object of discussion with their teachers and with one another, and this discussion further promotes the reflection on one's own thinking that is essential to good learning.

Thus self-assessment by pupils, far from being a luxury, is in fact *an essential component of formative assessment*. When anyone is trying to learn, feedback about the effort has three elements: recognition of the *desired goal*, evidence about *present position*, and some understanding of a *way to close the gap* between the two.15 All three must be understood to some degree by anyone before he or she can take action to improve learning.

Such an argument is consistent with more general ideas established by research into the way people learn. New understandings are not simply swallowed and stored in isolation; they have to be assimilated in relation to preexisting ideas. The new and the old may be inconsistent or even in conflict, and the disparities must be resolved by thoughtful actions on the part of the learner. Realizing that there are new goals for the learning is an essential part of this process of assimilation. Thus we conclude: *if formative assessment is to be productive, pupils should be trained in self-assessment so that they can understand the main purposes of their learning and thereby grasp what they need to do to achieve.*

The evolution of effective teaching. The research studies referred to above show very clearly that effective programs of formative assessment involve far more than the addition of a few observations and tests to an existing program. They require careful scrutiny of all the main components of a teaching plan. Indeed, it is clear that instruction and formative assessment are indivisible.

To begin at the beginning, the choice of tasks for classroom work and homework is important. Tasks have to be justified in terms of the learning aims that they serve, and they can work well only if opportunities for pupils to communicate their evolving understanding are built into the planning. Discussion, observation of activities, and marking of written work can all be used to provide those opportunities, but it is then important to look at or listen carefully to the talk, the writing, and the actions through which pupils develop and display the state of their understanding. Thus we maintain that *opportunities for pupils to express their understanding should be designed into any piece of teaching, for this will initiate the interaction through which formative assessment aids learning.*

Discussions in which pupils are led to talk about their understanding in their own ways are important aids to increasing knowledge and improving understanding. Dialogue with the teacher provides the opportunity for the teacher to respond to and reorient a pupil's thinking. However, there are clearly recorded examples of such discussions in which teachers have, quite unconsciously, responded in ways that would inhibit the future learning of a pupil. What the examples have in common is that the teacher is looking for a particular response and lacks the flexibility or the confidence to deal with the unexpected. So the teacher tries to direct the pupil toward giving the expected answer. In manipulating the dialogue in this way, the teacher seals off any unusual, often thoughtful

but unorthodox, attempts by pupils to work out their own answers. Over time the pupils get the message: they are not required to think out their own answers. The object of the exercise is to work out -- or guess -- what answer the teacher expects to see or hear.

A particular feature of the talk between teacher and pupils is the asking of questions by the teacher. This natural and direct way of checking on learning is often unproductive. One common problem is that, following a question, teachers do not wait long enough to allow pupils to think out their answers. When a teacher answers his or her own question after only two or three seconds and when a minute of silence is not tolerable, there is no possibility that a pupil can think out what to say.

There are then two consequences. One is that, because the only questions that can produce answers in such a short time are questions of fact, these predominate. The other is that pupils don't even try to think out a response. Because they know that the answer, followed by another question, will come along in a few seconds, there is no point in trying. It is also generally the case that only a few pupils in a class answer the teacher's questions. The rest then leave it to these few, knowing that they cannot respond as quickly and being unwilling to risk making mistakes in public. So the teacher, by lowering the level of questions and by accepting answers from a few, can keep the lesson going but is actually out of touch with the understanding of most of the class. The question/answer dialogue becomes a ritual, one in which thoughtful involvement suffers.

There are several ways to break this particular cycle. They involve giving pupils time to respond; asking them to discuss their thinking in pairs or in small groups, so that a respondent is speaking on behalf of others; giving pupils a choice between different possible answers and asking them to vote on the options; asking all of them to write down an answer and then reading out a selected few; and so on. What is essential is that any dialogue should evoke thoughtful reflection in which all pupils can be encouraged to take part, for only then can the formative process start to work. In short, the dialogue between pupils and a teacher should be *thoughtful, reflective, focused to evoke and explore understanding, and conducted so that all pupils have an opportunity to think and to express their ideas*.

Tests given in class and tests and other exercises assigned for homework are also important means of promoting feedback. A good test can be an occasion for learning. It is better to have frequent short tests than infrequent long ones. Any new learning should first be tested within about a week of a first encounter, but more frequent tests are counterproductive. The quality of the test items -- that is, their relevance to the main learning aims and their clear communication to the pupil -- requires scrutiny as well. Good questions are hard to generate, and teachers should collaborate and draw on outside sources to collect such questions.

Given questions of good quality, it is essential to ensure the quality of the feedback. Research studies have shown that, if pupils are given only marks or grades, they do not benefit from the feedback. The worst scenario is one in which some pupils who get low marks this time also got low marks last time and come to expect to get low marks next

time. This cycle of repeated failure becomes part of a shared belief between such students and their teacher. Feedback has been shown to improve learning when it gives each pupil specific guidance on strengths and weaknesses, preferably without any overall marks. Thus the way in which test results are reported to pupils so that they can identify their own strengths and weaknesses is critical. Pupils must be given the means and opportunities to work with evidence of their difficulties. For formative purposes, a test at the end of a unit or teaching module is pointless; it is too late to work with the results. We conclude that *the feedback on tests, seatwork, and homework should give each pupil guidance on how to improve, and each pupil must be given help and an opportunity to work on the improvement.*

All these points make clear that there is no one simple way to improve formative assessment. What is common to them is that a teacher's approach should start by being realistic and confronting the question "Do I really know enough about the understanding of my pupils to be able to help each of them?"

Much of the work teachers must do to make good use of formative assessment can give rise to difficulties. Some pupils will resist attempts to change accustomed routines, for any such change is uncomfortable, and emphasis on the challenge to think for yourself (and not just to work harder) can be threatening to many. Pupils cannot be expected to believe in the value of changes for their learning before they have experienced the benefits of such changes. Moreover, many of the initiatives that are needed take more class time, particularly when a central purpose is to change the outlook on learning and the working methods of pupils. Thus teachers have to take risks in the belief that such investment of time will yield rewards in the future, while "delivery" and "coverage" with poor understanding are pointless and can even be harmful.

Teachers must deal with two basic issues that are the source of many of the problems associated with changing to a system of formative assessment. The first is *the nature of each teacher's beliefs about learning*. If the teacher assumes that knowledge is to be transmitted and learned, that understanding will develop later, and that clarity of exposition accompanied by rewards for patient reception are the essentials of good teaching, then formative assessment is hardly necessary. However, most teachers accept the wealth of evidence that this transmission model does not work, even when judged by its own criteria, and so are willing to make a commitment to teaching through interaction. Formative assessment is an essential component of such instruction. We do not mean to imply that individualized, one-on-one teaching is the only solution; rather we mean that what is needed is a classroom culture of questioning and deep thinking, in which pupils learn from shared discussions with teachers and peers. What emerges very clearly here is the indivisibility of instruction and formative assessment practices.

The other issue that can create problems for teachers who wish to adopt an interactive model of teaching and learning relates to *the beliefs teachers hold about the potential of all their pupils for learning*. To sharpen the contrast by overstating it, there is on the one hand the "fixed I.Q." view -- a belief that each pupil has a fixed, inherited intelligence that cannot be altered much by schooling. On the other hand, there is the "untapped

potential" view -- a belief that starts from the assumption that so-called ability is a complex of skills that can be learned. Here, we argue for the underlying belief that all pupils can learn more effectively if one can clear away, by sensitive handling, the obstacles to learning, be they cognitive failures never diagnosed or damage to personal confidence or a combination of the two. Clearly the truth lies between these two extremes, but the evidence is that *ways of managing formative assessment that work with the assumptions of "untapped potential" do help all pupils to learn and can give particular help to those who have previously struggled.*

Policy and Practice

Changing the policy perspective. The assumptions that drive national and state policies for assessment have to be called into question. The promotion of testing as an important component for establishing a competitive market in education can be very harmful. The more recent shifting of emphasis toward setting targets for all, with assessment providing a touchstone to help check pupils' attainments, is a more mature position. However, we would argue that *there is a need now to move further, to focus on the inside of the "black box" and so to explore the potential of assessment to raise standards directly as an integral part of each pupil's learning work.*

It follows from this view that several changes are needed. First, policy ought to start with a recognition that the prime locus for raising standards is the classroom, so that the overarching priority has to be the promotion and support of change within the classroom. Attempts to raise standards by reforming the inputs to and measuring the outputs from the black box of the classroom can be helpful, but they are not adequate on their own. Indeed, their helpfulness can be judged only in light of their effects in classrooms.

The evidence we have presented here establishes that a clearly productive way to start implementing a classroom-focused policy would be to improve formative assessment. This same evidence also establishes that in doing so we would not be concentrating on some minor aspect of the business of teaching and learning. Rather, we would be concentrating on several essential elements: the quality of teacher/pupil interactions, the stimulus and help for pupils to take active responsibility for their own learning, the particular help needed to move pupils out of the trap of "low achievement," and the development of the habits necessary for all students to become lifelong learners. Improvements in formative assessment, which are within the reach of all teachers, can contribute substantially to raising standards in all these ways.

Four steps to implementation. If we accept the argument outlined above, what needs to be done? The proposals outlined below do not follow directly from our analysis of assessment research. They are consistent with its main findings, but they also call on more general sources for guidance.16

At one extreme, one might call for more research to find out how best to carry out such work; at the other, one might call for an immediate and large-scale program, with new guidelines that all teachers should put into practice. Neither of these alternatives is

sensible: while the first is unnecessary because enough is known from the results of research, the second would be unjustified because not enough is known about classroom practicalities in the context of any one country's schools.

Thus the improvement of formative assessment cannot be a simple matter. There is no quick fix that can alter existing practice by promising rapid rewards. On the contrary, if the substantial rewards promised by the research evidence are to be secured, each teacher must find his or her own ways of incorporating the lessons and ideas set out above into his or her own patterns of classroom work and into the cultural norms and expectations of a particular school community.17 This process is a relatively slow one and takes place through sustained programs of professional development and support. This fact does not weaken the message here; indeed, it should be seen as a sign of its authenticity, for lasting and fundamental improvements in teaching and learning must take place in this way. A recent international study of innovation and change in education, encompassing 23 projects in 13 member countries of the Organisation for Economic Co-operation and Development, has arrived at exactly the same conclusion with regard to effective policies for change.18 Such arguments lead us to propose a four-point scheme for teacher development.

1. *Learning from development.* Teachers will not take up ideas that sound attractive, no matter how extensive the research base, if the ideas are presented as general principles that leave the task of translating them into everyday practice entirely up to the teachers. Their classroom lives are too busy and too fragile for all but an outstanding few to undertake such work. What teachers need is a variety of living examples of implementation, as practiced by teachers with whom they can identify and from whom they can derive the confidence that they can do better. They need to see examples of what doing better means in practice.

So changing teachers' practice cannot begin with an extensive program of training for all; that could be justified only if it could be claimed that we have enough "trainers" who know what to do, which is certainly not the case. The essential first step is to set up a small number of local groups of schools -- some primary, some secondary, some inner-city, some from outer suburbs, some rural -- with each school committed both to a school-based development of formative assessment and to collaboration with other schools in its local group. In such a process, the teachers in their classrooms will be working out the answers to many of the practical questions that the evidence presented here cannot answer. They will be reformulating the issues, perhaps in relation to fundamental insights and certainly in terms that make sense to their peers in other classrooms. It is also essential to carry out such development in a range of subject areas, for the research in mathematics education is significantly different from that in language, which is different again from that in the creative arts.

The schools involved would need extra support in order to give their teachers time to plan the initiative in light of existing evidence, to reflect on their experience as it develops, and to offer advice about training others in the future. In addition, there would be a need for external evaluators to help the teachers with their development work and to collect

evidence of its effectiveness. Video studies of classroom work would be essential for disseminating findings to others.

2. *Dissemination.* This dimension of the implementation would be in low gear at the outset -- offering schools no more than general encouragement and explanation of some of the relevant evidence that they might consider in light of their existing practices. Dissemination efforts would become more active as results and resources became available from the development program. Then strategies for wider dissemination -- for example, earmarking funds for inservice training programs -- would have to be pursued.

We must emphasize that this process will inevitably be a slow one. To repeat what we said above, *if the substantial rewards promised by the evidence are to be secured, each teacher must find his or her own ways of incorporating the lessons and ideas that are set out above into his or her own patterns of classroom work.* Even with optimum training and support, such a process will take time.

3. *Reducing obstacles.* All features in the education system that actually obstruct the development of effective formative assessment should be examined to see how their negative effects can be reduced. Consider the conclusions from a study of teachers of English in U.S. secondary schools.

Most of the teachers in this study were caught in conflicts among belief systems and institutional structures, agendas, and values. The point of friction among these conflicts was assessment, which was associated with very powerful feelings of being overwhelmed, and of insecurity, guilt, frustration, and anger. . . . This study suggests that assessment, as it occurs in schools, is far from a merely technical problem. Rather, it is deeply social and personal.19

The chief negative influence here is that of short external tests. Such tests can dominate teachers' work, and, insofar as they encourage drilling to produce right answers to short, out-of-context questions, they can lead teachers to act against their own better judgment about the best ways to develop the learning of their pupils. This is not to argue that all such tests are unhelpful. Indeed, they have an important role to play in securing public confidence in the accountability of schools. For the immediate future, what is needed in any development program for formative assessment is to study the interactions between these external tests and formative assessments to see how the models of assessment that external tests can provide could be made more helpful.

All teachers have to undertake some summative assessment. They must report to parents and produce end-of-year reports as classes are due to move on to new teachers. However, the task of assessing pupils summatively for external purposes is clearly different from the task of assessing ongoing work to monitor and improve progress. Some argue that these two roles are so different that they should be kept apart. We do not see how this can be done, given that teachers must have some share of responsibility for the former and must take the leading responsibility for the latter.20 However, teachers clearly face

difficult problems in reconciling their formative and summative roles, and confusion in teachers' minds between these roles can impede the improvement of practice.

The arguments here could be taken much further to make the case that teachers should play a far greater role in contributing to summative assessments for accountability. One strong reason for giving teachers a greater role is that they have access to the performance of their pupils in a variety of contexts and over extended periods of time.

This is an important advantage because sampling pupils' achievement by means of short exercises taken under the conditions of formal testing is fraught with dangers. It is now clear that performance in any task varies with the context in which it is presented. Thus some pupils who seem incompetent in tackling a problem under test conditions can look quite different in the more realistic conditions of an everyday encounter with an equivalent problem. Indeed, the conditions under which formal tests are taken threaten validity because they are quite unlike those of everyday performance. An outstanding example here is that collaborative work is very important in everyday life but is forbidden by current norms of formal testing.21 These points open up wider arguments about assessment systems as a whole -- arguments that are beyond the scope of this article.

4. *Research.* It is not difficult to set out a list of questions that would justify further research in this area. Although there are many and varied reports of successful innovations, they generally fail to give clear accounts of one or another of the important details. For example, they are often silent about the actual classroom methods used, the motivation and experience of the teachers, the nature of the tests used as measures of success, or the outlooks and expectations of the pupils involved.

However, while there is ample justification for proceeding with carefully formulated projects, we do not suggest that everyone else should wait for their conclusions. Enough is known to provide a basis for active development work, and some of the most important questions can be answered only through a program of practical implementation.

Directions for future research could include a study of the ways in which teachers understand and deal with the relationship between their formative and summative roles or a comparative study of the predictive validity of teachers' summative assessments versus external test results. Many more questions could be formulated, and it is important for future development that some of these problems be tackled by basic research. At the same time, experienced researchers would also have a vital role to play in the evaluation of the development programs we have proposed.

Are We Serious About Raising Standards?

The findings summarized above and the program we have outlined have implications for a variety of responsible agencies. However, it is the responsibility of governments to take the lead. It would be premature and out of order for us to try to consider the relative roles in such an effort, although success would clearly depend on cooperation among government agencies, academic researchers, and school-based educators.

The main plank of our argument is that standards can be raised only by changes that are put into direct effect by teachers and pupils in classrooms. There is a body of firm evidence that formative assessment is an essential component of classroom work and that its development can raise standards of achievement. We know of no other way of raising standards for which such a strong prima facie case can be made. Our plea is that national and state policy makers will grasp this opportunity and take the lead in this direction.

---

1. James W. Stigler and James Hiebert, "Understanding and Improving Classroom Mathematics Instruction: An Overview of the TIMSS Video Study," *Phi Delta Kappan*, September 1997, pp. 19-20.
2. There is no internationally agreed-upon term here. "Classroom evaluation," "classroom assessment," "internal assessment," "instructional assessment," and "student assessment" have been used by different authors, and some of these terms have different meanings in different texts.
3. Paul Black and Dylan Wiliam, "Assessment and Classroom Learning," *Assessment in Education*, March 1998, pp. 7-74.
4. Lynn S. Fuchs and Douglas Fuchs, "Effects of Systematic Formative Evaluation: A Meta-Analysis," *Exceptional Children*, vol. 53, 1986, pp. 199-208.
5. See Albert E. Beaton et al., *Mathematics Achievement in the Middle School Years* (Boston: Boston College, 1996).
6. Lynn S. Fuchs et al., "Effects of Task-Focused Goals on Low-Achieving Students with and Without Learning Disabilities," *American Educational Research Journal*, vol. 34, 1997, pp. 513-43.
7. OFSTED (Office for Standards in Education), *Subjects and Standards: Issues for School Development Arising from OFSTED Inspection Findings 1994-5: Key Stages 3 and 4 and Post-16* (London: Her Majesty's Stationery Office, 1996), p. 40.
8. Nicholas Daws and Birendra Singh, "Formative Assessment: To What Extent Is Its Potential to Enhance Pupils' Science Being Realized?," *School Science Review*, vol. 77, 1996, p. 99.
9. Clement Dassa, Jesús Vazquez-Abad, and Djavid Ajar, "Formative Assessment in a Classroom Setting: From Practice to Computer Innovations," *Alberta Journal of Educational Research*, vol. 39, 1993, p. 116.
10. D. Monty Neill, "Transforming Student Assessment," *Phi Delta Kappan*, September 1997, pp. 35-36.
11. *Task Group on Assessment and Testing: A Report* (London: Department of Education and Science and the Welsh Office, 1988).
12. Richard Daugherty, *National Curriculum Assessment: A Review of Policy, 1987-1994* (London: Falmer Press, 1995).
13. Terry A. Russell, Anne Qualter, and Linda McGuigan, "Reflections on the Implementation of National Curriculum Science Policy for the 5-14 Age Range: Findings and Interpretations from a National Evaluation Study in England," *International Journal of Science Education*, vol. 17, 1995, pp. 481-92.
14. Phillipe Perrenoud, "Towards a Pragmatic Approach to Formative Evaluation," in Penelope Weston, ed., *Assessment of Pupils' Achievement: Motivation and School Success* (Amsterdam: Swets and Zeitlinger, 1991), p. 92.

15. D. Royce Sadler, "Formative Assessment and the Design of Instructional Systems," *Instructional Science*, vol. 18, 1989, pp. 119-44.
16. Paul J. Black and J. Myron Atkin, *Changing the Subject: Innovations in Science, Mathematics, and Technology Education* (London: Routledge for the Organisation for Economic Co-operation and Development, 1996); and Michael G. Fullan, with Suzanne Stiegelbauer, *The New Meaning of Educational Change* (London: Cassell, 1991).
17. See Stigler and Hiebert, pp. 19-20.
18. Black and Atkin, op. cit.
19. Peter Johnston et al., "Assessment of Teaching and Learning in Literature-Based Classrooms," *Teaching and Teacher Education*, vol. 11, 1995, p. 359.
20. Dylan Wiliam and Paul Black, "Meanings and Consequences: A Basis for Distinguishing Formative and Summative Functions of Assessment," *British Educational Research Journal*, vol. 22, 1996, pp. 537-48.
21. These points are developed in some detail in Sam Wineburg, "T. S. Eliot, Collaboration, and the Quandaries of Assessment in a Rapidly Changing World," *Phi Delta Kappan*, September 1997, pp. 59-65.

PAUL BLACK is professor emeritus in the School of Education, King's College, London, where DYLAN WILIAM is head of school and professor of educational assessment.