

The Role of Interim Formative Assessment in Student Achievement

Dr Linda Griffith - University of Central Arkansas

Ms Lisa Bailey - Southeast Arkansas Educational Service Cooperative

Mr Tony Timms – Southeast Arkansas Educational Service Cooperative

IAEA 2008

Abstract

This paper reports on an existing research project in which interim formative assessments were implemented as a component of a five-step process for improving student learning. This five-step process includes vertical and horizontal curriculum alignment, instructional alignment, suggested order of curriculum, interim formative assessments, and use of data to monitor and modify curriculum and instruction. The work of Wiliam and Stiggins established the foundation for this project, titled Target Assessment. The Target Assessment project is a collaborative effort of 13 regional educational service cooperatives in the state of Arkansas, more than 150 Arkansas school districts, the University of Central Arkansas, the Arkansas Department of Education, and a private technology company – Triand. This paper describes the mathematics, literacy, and science components of this assessment project, the validation processes for developing the item banks, and the ongoing reliability studies. The implementation of interim formative assessment in typical participating school districts is explained, and the evidence of student learning progress in these districts is also presented.

The Role of Interim Formative Assessment in Student Achievement

The federal government of the United States wishes to ensure that all children have a fair, equal, and significant opportunity to obtain a high-quality education and reach, at a minimum, proficiency on challenging State academic standards so that students, teachers, parents, and administrators can measure progress against common expectations for student academic achievement. This has resulted in mandated end-of-year high-stakes assessments that are aligned with state learning standards. However, these end-of-year assessments are not designed to provide decision-makers with immediate specific feedback with which to improve student learning in individual classrooms. While it is important to raise the standards for student learning; the task of improving student learning is much more complex.

The work of Richard Stiggins (1992) and Black and Wiliam (1998) established the foundation for the Interim Target Formative Assessment Project. Based on research, this project has identified a 5-step process for improving student learning:

- 1) Alignment of curriculum and instruction both vertically (grade to grade) and horizontally (across teachers at the same grade).
- 2) Alignment of instruction including selection of instructional materials, participation in professional development on effective instructional strategies, and the use of minute-by-minute, day-by-day and week-by-week formative assessment in the classroom.
- 3) Creation of pacing guides or curriculum maps that articulate the plan of instruction for the year.
- 4) Administration of interim formative assessments.
- 5) Use the data from interim formative assessments to monitor and modify the alignment, instruction, and pacing.

Learning occurs because of what teachers and students do in classrooms. Frequent feedback contributes to gains in student learning. The most helpful feedback, producing the greatest gains, is that which helps students become aware of gaps that exist between the learning goal and current knowledge, understanding or skill through specific, carefully-focused feedback (Bangert-Drowns, Kulick & Morgan, 1991; Ramaprasad, 1983; Sadler, 1989) Students who have an understanding of their learning goals and assessment strategies, and have opportunities to discuss and reflect on their work, show greater progress than those who do not (Fontana & Fernandes, 1994; Frederickson & White, 1997). Interim formative assessments inform the student and the teacher about individual and class progress. This information can be used to improve student achievement beyond a particular classroom.

Black and Wiliam (1998) define formative assessment as . . . “all those activities undertaken by teachers, and by the students in assessing themselves, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged. Such assessment becomes formative when the evidence is actually used to adapt the teaching to meet student needs.” (p.2) As this research indicates, a variety of assessments are used by teachers throughout the school year to make decisions about curriculum, instruction, and student needs; these assessments provide teachers with information with which they can determine whether to re-teach a lesson, move forward, or provide small group or individual instruction. We recognize that this type of formative assessment occurs in 3 ways:

- 1- Immediate - Teachers instruct and use questioning strategies and classroom assessments to determine if changes in instruction are needed.

- 2- End-of-year, High-Stakes – These assessments provide administrators and teachers access to specific data, but the information is received too late in the year for adjustments to be made in instruction.
- 3- Interim - This type of formal formative assessment provides data for students, teachers, and administrators throughout the year, allowing them to monitor student progress and modify instruction in a timely manner.

Although each of these assessments provides valuable information, only interim formative assessments offer insight for decision-makers into the curriculum, instruction, and progress for individual students on a continuous basis. The quality of instruction is dependent on this kind of information. Interim formative assessments provide valuable feedback to the following stakeholders:

Students - Interim formative assessments give students feedback regarding the progress they have made toward meeting the goals set by high-stakes tests, identifying areas in which additional learning is needed and providing them with the opportunity to gain knowledge in these areas prior to the administration of the high-stakes test.

Parents - Interim formative assessments help parents to understand expectations and assessment criteria. Parents can use data from interim formative assessments to monitor their child's progress throughout the year; in this way, they become an advocate for the child by making sure the child's specific learning needs are addressed.

Teachers - Interim formative assessments give teachers a chance to reflect on their practice up to this point. These assessments help teachers to answer questions like: Did the developer of this test interpret the student learner expectations in the same way I did? Can students transfer the context in which something was taught to the context it is asked on the assessment? Has the instructional vocabulary that is used in the classroom been adequate? Has my instruction transferred to an assessment that is written by someone else?

Administrators - Administrators are not privy to the day-by-day feedback that teachers gather in their classrooms, and by the time they get the results of the high-stakes tests, it's too late to make needed adjustments. Therefore, interim formative assessments allow administrators to monitor student progress and make modifications more frequently. The assessments provide data to answer questions like: Have the instructional materials we've provided been adequate? Do we need to provide alternative instructional materials? What are our immediate professional development needs? Do all faculty members need the same professional development or is there a need for individualization of professional development? Is the current curriculum and pacing appropriate?

Classroom grades often do not correlate to the end-of-year test scores. One reason for this is that classroom summative assessments are written and given within the same context and environment as the teacher's instruction. Teachers assess what they taught, which may or may not be what the high-stakes tests will assess. It is because of the advantages of interim formative assessments listed above that thirteen of Arkansas's educational cooperatives felt the need to offer the school districts they serve the opportunity to participate in such an endeavour. The interim target assessment project has three goals:

1. To provide students exposure to context and vocabulary that are not necessarily matched with the teacher's instructional context and vocabulary. Interim assessments offer students an occasion to apply what has been taught in the classroom. This type of assessment helps students and teachers to determine whether students can transfer what has been taught. It helps students to understand "What was I supposed to know?" and "Do I know it?" Students partner with their teacher and continually monitor their current

- level of attainment in relation to specified student learning expectations. In this way, students can set learning goals and take ownership of the learning.
2. To provide teachers with access to additional data from assessments written by someone else based on the published learning goals. These assessments are more formal than classroom assessments and help prepare students for the high-stakes testing format. This data also gives teachers an opportunity to be reflective practitioners based on student performance. With this type of assessment, the teacher can reflect on whether the interpretation of the state standard may have been different, or the context or vocabulary on the assessment may have been different from his or her instruction.
 3. To provide school and district decision-makers with recurrent information regarding student performance so that they can better monitor and modify to meet student needs. Monitoring student performance closely throughout the year, allows administrators to make better decisions about appropriate professional development, curriculum needs, and individual student progress.

The results of each formative assessment should be used exclusively to guide future learning. Therefore, teachers are discouraged from giving a grade to these assessments, thereby making them summative. By not assigning a grade, these assessments help students to become self-monitoring; not only checking their progress, but making a plan to improve their progress before the next interim assessment. According to Richard Stiggins (1992), “The benefits of clear targets to students are indisputable . . . the key feature to student success is students knowing where they are going, that is, understanding what they are to learn.” By using assessments in this manner, students are empowered to take charge of their own learning.

The Interim Target Formative Assessment Project is a collaborative effort of 13 regional educational service cooperatives representing more than 150 school districts in Arkansas, a university (University of Central Arkansas), the Arkansas Department of Education, and a private technology company, Triand. The 13 educational cooperatives fund the project, provide third round proofreaders for the assessments and scan the multiple-choice answer sheets for the school districts. Funding from the co-ops provides 50% release time for a university professor to lead the project and provides a mathematics director and literacy director. Their well-trained staff also provides professional development to participating schools. Teachers are trained to access and analyze data. Individual schools within each co-op area copy and administer the assessments.

The University of Central Arkansas supervises the project design, development of the tests, and provides professional development in collaboration with the education service cooperatives. It is also responsible for conducting the reliability studies and validation of items.

The Arkansas Department of Education (ADE) provides access to the software that allows the test to be scanned and the data to be analyzed. ADE supports the project by offering specific professional development and training for the specialists in the co-ops so that they may assist the schools in using the data.

Triand, a private technology company, provides software for test development, scanning and data analysis. Triand offers a secure, web-based software application that helps parents, teachers, and school districts organize and track a student's progress. The goal of Triand is to assist educational leaders in meeting the challenges of a standards-based educational environment with features that give decisional support designed to improve student performance. Triand has the only collaborative resource sharing in the nation. Users are able to share lesson plans, instructional resources, assessments items aligned to state standards at the click of a button.

With these support systems in place, the focus for the first year of the project was on mathematics and literacy in grades 3-11. In year two, we continued the focus in these areas, and piloted science assessments in grades 5, 7 and biology. In year three, 2008-2009, we will provide interim formative assessments in literacy and mathematics for grades 3-11, and in science for grades 5, 7 and biology. The education service cooperatives and the school districts have created curriculum documents based on the work of Lisa Carter (Total Instructional Alignment) that articulate what is to be taught, when it is to be taught, and suggest appropriate resources.

The University of Central Arkansas has supervised the development of five tests at each grade. A pre-test, to be given in the early part of the school year, provides students with an opportunity to see a picture of what the expected outcome of that school year in that subject will be. On this test, in 2008-2009, students will be given 5 options, a correct answer, three distracters, and "I don't know yet". This pre-test will permit students to evaluate what they already know and to determine their learning goals for the year.

The pre-test provides teachers with information regarding their class. This information gives teachers a chance to make modifications in the pacing guide by reducing instructional time in the areas in which their students demonstrated strength, and increasing instructional time in the areas of greatest need. The pre-test data provides building and district administrators an opportunity to see the strengths and/or weaknesses in their vertical alignment. In the Arkansas Frameworks there is a considerable amount of overlap from grade to grade in regard to student learning expectations (SLEs). Therefore, a good pre-test result indicates a solid vertical alignment, and poor pre-test results may indicate a need to strengthen the vertical alignment.

The second-fourth assessments provided by this project are interim formative assessments. Although we feel that three formative assessments per year are not really enough, given the budget and time constraints we have, we are limited to only three. These three assessments are given at roughly equal intervals between the beginning of school and the time in which the high-stakes state summative assessments are given (in the 8th month of the school year.)

The fifth assessment, a post-test is provided at the end of the year. In mathematics, science, and grammar, the pre- and post-test items assess the same concepts so that an item by item comparison can be made by students. In reading, this policy is followed as closely as possible; however, some variations are necessary because different passages provide opportunities to assess different concepts. The purpose of the post-test is primarily to allow students to see the progress since the beginning of the year. What do I know now that I did not know at the beginning of this year? For teachers and administrators, pre- and post- tests are helpful when doing year-end curriculum audits. On the concepts where little or no growth is indicated, changes need to be planned for next year.

Teachers are discouraged from looking at an average or percentage grade on these assessments, but are encouraged to examine the results item-by-item. This will help them determine if the writer of the test interpreted the SLE the same way they did. Also they need to determine if they have not taught an SLE that is tested because they are behind in the planned pacing of instruction. For each item, the teachers should do the following:

- Determine if 70% or more (may be higher for high performing schools) of the students were successful on an item or set of items assessing a particular SLE. In this case the teachers should mark their plans, pacing guides, and materials, in some way to recognize that this instruction was successful and should be use it in subsequent years.

- Determine if 30% or less (may be higher for high performing schools) of the students were successful on an item or set of items assessing a particular SLE. In this case the teachers should plan to re-teach this concept to the whole group using different materials and strategies. Support personnel such as district or regional specialists may be contacted for assistance in selecting these materials and strategies.
- For the remaining items, ones for which results were mediocre, time may not allow for re-teaching these whole group. However, for next year, teachers should try to improve their instructional plans in these areas.
- For each item the teachers need to determine if the same distracter answer was selected by a large group of students. This may indicate that a misconception has developed and this needs to be addressed as soon as possible.
- Finally, classroom teachers, and those charged with remediation responsibility, should look at individual student results to see areas in which individual students need additional instruction/remediation. This remediation should take place in small groups, individually, or beyond the regular classroom.

Administrators examine each set of interim formative assessment data by determining the following:

- If the performance on an item is consistently good among all the teachers in this grade/level. If so they may conclude that instructional materials, professional development, and time within the pacing guide were adequate.
- If an item is consistently poor, across a single grade-level, then additional research must be done to determine the reason for this poor performance. Do we need different instructional materials? Was the pacing guide unrealistic? Do we need to provide specific professional development for teachers?
- If there are items where the results are not consistent from teacher to teacher within a grade/level, this indicates horizontal alignment issues. These issues could be due to the lack of knowledge of a novice teacher or due to teachers that were unable to attend a professional development, or because of teachers having different levels of content knowledge based on pre-service training. In these situations, teachers that are experiencing success can serve as resources for struggling teachers. Collaboration, not competition, among grade level teams is encouraged.

When creating interim formative assessments and pre- and post-tests, the project has worked extensively to be sure that the items are valid and the assessments are reliable. Each item is validated using the following procedure: An expert gives his/her opinion on the degree to which the item tests the SLE that it purports to test. After the test has been constructed, paid proofreaders, who are master teachers, examine the entire test suggesting changes and giving their opinion on the validity of each item. Finally, literacy, mathematics, and science specialists at the educational cooperatives proof the tests and give their opinion on the validity of each item.

Reliability studies are correlating student performance on the interim formative assessments with the results of the high-stakes state tests. This paper includes the results for two schools one from year one of the project and one for year two. Much more extensive correlations are being undertaken and the results will be reported when those are completed. In the first year of the project the literacy data that has been examined shows correlation coefficients ranging from 0.63 to 0.77 for the individual formative assessments correlated with the high stakes test results. A composite score of all three formative assessments correlated with the high-stakes test results had

r-values of 0.80 and 0.81. The correlations with the post-test given in year one for literacy ranged from 0.72 to 0.76. These results are shown in Table 1 below.

In the first year of the project the mathematics data that has been examined shows correlation coefficients ranging from 0.10 to 0.80 for the individual formative assessments correlated with the high stakes test results. There was an obvious reliability issue with the first formative assessment given at grade three. A composite score of all three formative assessments correlated with the high-stakes test results had r-values ranging from 0.75 and 0.85. The correlations with the post-test given in year one for mathematics ranged from 0.74 to 0.83. These results are shown in Table 3 below.

The prediction equation for the composite formative assessment was used to predict the percentage of students that would score proficient of the high-stakes test. Tables 2 and 4 show these results. All the predicted percentages were within 10 percentage points of the actual results. Finally, predictions for each individual student were compared with the student's actual performance. Results varied by grade level and ranges from 54% to 78%.

Table 1: 2006-2007 Literacy Reliability Correlation Data for School A

Grade	Sample Size	R value Formative 1	R Value Formative 2	R Value Formative 3	R Value Composite	R Value Post-test
5	107	0.77	0.73	0.65	0.80	0.72
6	106	0.75	0.77	0.73	0.81	0.76
7	139	0.63	0.66	0.68	0.81	0.73

Table 2: 2006-2007 Literacy Predicted Results vs. Actual Results School A

Grade	Predicted Passing Percentage	Actual Passing Percentage	Percentage Predicted Correctly
5	87%	77%	78%
6	58%	62%	64%
7	61%	55%	65%

Table 3: 2006-2007 Mathematics Reliability Correlation Data for School A

Grade	Sample Size	R value Formative 1	R Value Formative 2	R Value Formative 3	R Value Composite	R Value Post-test
3	126	0.10	0.71	0.75	0.75	0.74
4	120	0.63	0.68	0.74	0.81	0.82
5	99	0.73	0.76	0.67	0.83	0.77
6	110	0.61	0.74	0.75	0.81	0.80
7	129	0.68	0.78	0.80	0.85	0.83

Table 4: 2006-2007 Mathematics Predicted & Actual Results School A

Grade	Predicted Passing Percentage	Actual Passing Percentage	Percentage Predicted Correctly
3	55%	55%	78%
4	64%	66%	54%
5	82%	76%	70%
6	75%	69%	65%
7	71%	65%	73%

In the second year of the project the literacy data that has been examined shows correlation coefficients ranging from 0.62 to 0.79 for the individual formative assessments correlated with the high stakes test results. A composite score of all three formative assessments correlated with the high-stakes test results had r-values ranging from 0.72 to 0.83. The correlations with the post-test given in year one for literacy ranged from 0.72 to 0.82. These results are shown in Table 5 below.

In the second year of the project the mathematics data that has been examined shows correlation coefficients ranging from 0.58 to 0.77 for the individual formative assessments correlated with the high stakes test results. A composite score of all three formative assessments correlated with the high-stakes test results had r-values ranging from 0.74 and 0.85. The correlations with the post-test given in year one for mathematics ranged from 0.72 to 0.83. These results are shown in Table 7 below.

The prediction equation for the composite formative assessment was used to predict the percentage of students that would score proficient of the high-stakes test. Tables 6 and 8 show these results. All the predicted percentages were within 6 percentage points of the actual results. Finally, predictions for each individual student were compared with the student’s actual performance. Results varied by grade level and ranges from 77% to 91%.

In year two there was a pre-test and growth from the pre-test to the post-test was computed. This growth ranged from 6 to 34 percentage points. The lack of significant change in pre- and post test results at some grade level was surprising.

Table 5: 2007-2008 Literacy Reliability Correlation Data for School B

Grade	Sample Size	R value Formative 1	R Value Formative 2	R Value Formative 3	R Value Composite	R Value Post-test
3	161	0.74	0.67	0.62	0.79	0.72
4	157	0.72	0.63	0.64	0.80	0.81
5	162	0.72	0.70	0.66	0.79	0.73
6	141	0.79	0.72	0.63	0.83	0.82

Table 6: 2007-2008 Literacy Predicted Results vs. Actual Results & Growth School B

Grade	Predicted Passing Percentage	Actual Passing Percentage	Percentage Predicted Correctly	Percentage Points Growth from Pre-test to Post-test
3	77%	78%	81%	24
4	74%	79%	86%	34
5	77%	81%	87%	6
6	82%	78%	91%	13

Table 7: 2006-2007 Mathematics Reliability Correlation Data for School B

Grade	Sample Size	R value Formative 1	R Value Formative 2	R Value Formative 3	R Value Composite	R Value Post-test
3	171	0.71	0.66	0.70	0.80	0.77
4	151	0.70	0.69	0.70	0.80	0.75
5	162	0.58	0.60	0.65	0.74	0.72
6	146	0.74	0.70	0.77	0.85	0.83

Table 8: 2006-2007 Mathematics Predicted vs. Actual Results & Growth School B

Grade	Predicted Passing Percentage	Actual Passing Percentage	Percentage Predicted Correctly	Percentage Points Growth from Pre-test to Post-test
3	92%	86%	88%	15
4	80%	83%	87%	19
5	83%	77%	77%	16
6	86%	80%	80%	16

The goals of this project for the future are to continue to increase the reliability of the interim formative assessments and continue to help students, teachers and administrators to use this data to improve student achievement.

References

- Banger-Drowns, R.L., Kulich, J.A., & Morgan, M.T. (1991). The Instructional Effect of Feedback in Test-Like Events. *Review of Educational Research*, 61, 213-238.
- Black, P.J. & Wiliam, D. (1998). Inside the Black Box: Raising the Standards Through Classroom Assessment. *Phi Delta Kappan*, (online), Retrieved 7/7/08 <http://www.pdkintl.org/kappan/kbla9810.html>
- Fontana, D., & Fernandes, M. (1994). Improvements in Mathematics Performance as a Consequence of Self-Assessment in Portugese Primary School Pupils. *British Journal of Educational Psychology*, 64(3), 407-414.
- Frederickson, J.R. & White, B.Y. (1997). Reflective Assessment of Students' Research within an Inquiry-Based Middle School Science Curriculum. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Ramaprasad, A. (1983). On the Definition of Feedback. *Behavioral Science*, 21(1), 4-13.
- Sadler, D.R. (1989). Formative Assessment and the Design of Instructional Systems. *Instructional Science*, 18, pp. 119-144.
- Stiggins, R. (1992). R. & Conklin, N.F. (1992) In *Teacher's Hands: Investigating the Practices of Classroom Assessment*. Albany, NY: SUNY Press.
- Stiggins, R. (1992). High Quality Classroom Assessment: What Does it Really Mean? NCME Instructional Topics in Educational Measurement Series, Module 12, summer 1992. (online), Retrieved 7/7/08 <http://www.ncme.org/pubs/items/19>