



**ALBUQUERQUE  
PUBLIC SCHOOLS**

# **Assess 2 Learn: Use & Effects of an Interim Benchmark Assessment**

Debra Heath, MPH  
Nancy Carrillo, Ph.D  
October 2008



# ALBUQUERQUE PUBLIC SCHOOLS

## BOARD OF EDUCATION

MARY LEE MARTIN  
President

BERNA FACIO  
Vice President

DOLORES GRIEGO  
Secretary

JON BARELA  
Instructional Policy Chair

PAULA MAES  
District Relations Chair

ROBERT D. LUCERO  
Finance/Audit Chair

MARTIN R. ESQUIVEL  
Capital Outlay Chair

## Superintendent WINSTON BROOKS

LINDA SINK  
Chief Academic Officer

BRAD WINTER  
Chief Operations Officer

RUBY ETHRIDGE  
Associate Superintendent

DIANE KERSCHEN  
Associate Superintendent

RAQUEL REEDY  
Associate Superintendent

EDUARDO SOTO  
Associate Superintendent

## INSTRUCTION AND ACCOUNTABILITY

Rose-Ann McKernan, Executive Director

### Research, Deployment & Accountability

Thomas Genné, Director

6400 Uptown Blvd. NE (400 EAST)

Albuquerque, New Mexico 87110

(505) 872-6870

[www.rda.aps.edu](http://www.rda.aps.edu)

## Executive Summary

APS adopted Assess2Learn (A2L), an on-line assessment program, to provide interim benchmark assessments of student progress toward state and district standards in reading and math three times a year. Purveyors of A2L theorize that when schools do a good job of implementing A2L, teachers are more likely to use A2L to inform instruction, and that student achievement should improve as a result. Anecdotal information suggests that A2L implementation varies across the district. This study seeks to address the following research questions:

1. Do teachers who use A2L results have higher levels of student achievement than teachers who do not use A2L results?
2. What are the nature, frequency and quality of A2L use by APS teachers? What are the main barriers to greater use?
3. Do teachers who use A2L results perceive that, as a result of A2L, their instruction has improved?

RDA used a ten-item survey to collect data from 4<sup>th</sup> through 8<sup>th</sup> grade teachers of reading and math on A2L use and perceived benefits and barriers. A total of 777 teachers completed usable surveys during district professional development workshops in September 2007, for an estimated response rate of 44 percent. Of the 777 respondents, 403 teachers were linked with students for outcome analysis.

Survey results provide some evidence that, in middle school, students whose teachers compare A2L results over time score better on their spring A2L assessments than students whose teachers do not. More detailed A2L reports and the frequency of use may impact scores, though these results are mixed. There is no evidence that the ways A2L results are used – for example to conference with parents, group students, or identify gaps in knowledge – impact student achievement, perhaps because teachers can perform these functions using other sources of information.

On the other hand, almost all teachers who use A2L results at the highest level agree that A2L improves their instructional practice in at least one way. Most agree that it helps them tailor instruction and address a wide range of learning needs. Seventy percent say it helps them improve their overall instructional effectiveness.

Survey results suggest that most teachers obtain A2L results for every test window. Teachers are most likely to obtain the Class List report, which lists overall test scores by student. Almost half obtain one or more detailed reports.

Most survey respondents report using A2L results in some way, but few use A2L to help modify classroom practices. At the most basic level, 77% of respondents say they discuss results with colleagues, 61% compare their class's fall and winter A2L results, and 57% rely on A2L to determine AIP eligibility. About half of teachers report using A2L results to inform their instructional practice. However, no more than 16% qualify as "high users," teachers who use A2L reports to plan what to teach, plus at least two of these other purposes: to group students, to identify gaps in knowledge/understanding, and/or to reflect on the effectiveness of teaching practices. Very few administer diagnostic testlets, the kind of classroom formative assessment that research has linked with improved student achievement.

Survey results suggest that elementary teachers are more likely to use A2L results than middle school teachers. Among middle school teachers, math teachers are more likely to use results than reading teachers.

Technical difficulties is the most commonly cited top barrier to using A2L results more, followed by difficulty in accessing reports, not having enough time, and not knowing how to access reports.

Several study limitations hinder our ability to speak with confidence about results and conclusions. We did not collect information on the depth, frequency or breadth of use. Self-reported data are inherently limited by biases. The sample of middle school reading teachers is inexplicably small. Most importantly, high/proficient users of A2L are likely to be high/proficient users of other assessment and instructional processes. Since this study did not collect data on other aspects of instruction, it is suggestive but not definitive that A2L use contributed to a good teacher's overall effectiveness.

The literature on classroom formative assessment provides compelling evidence that teachers who integrate assessment into their everyday instructional practice do, indeed, help their students achieve significant learning gains. A2L survey tests fit into the category of interim benchmark assessments, for which no research has demonstrated short term academic benefit. A2L diagnostic testlets, however, may be a powerful tool for advancing student achievement -- if teachers use them to actively engage students in assessing progress toward learning targets, and if they provide descriptive feedback and guidance about exactly what steps students can take to improve.

## Background

Assess2Learn (A2L) is an on-line assessment program produced by Riverside Publishing Company. APS adopted A2L in 2003-04, just prior to state legislation requiring every school district to have a 'short-cycle' assessment in 4<sup>th</sup> through 8<sup>th</sup> grade reading and math. Survey tests are mandated for grades 4 through 8 during specific testing windows three times per year (fall, winter, spring). A2L is designed to provide interim benchmark assessments of student progress toward state and district standards in reading and math.

Hundreds of A2L diagnostic 'testlets' are available also. These allow teachers to conduct deeper and/or more frequent assessments of student understanding. Alternatively, teachers may draw from an item bank or use their own items to construct custom assessments.

A2L assessment reports are available in a variety of formats (item analysis, skill proficiency, skill gap, summary scores) for entire class lists and for individual students. Schools must use A2L survey test results to determine students' eligibility for Academic Improvement Plans and summer school. District training materials also encourage teachers to use A2L for ongoing instructional decision-making.

The APS Teachers Quick Reference Guide to Assessment (RDA, January 2008) places A2L survey assessments in the categories of summative and benchmark. As defined by the APS Standards-Based Education (SBE) Implementation Plan (January 2007), these are *assessments of learning*. The Quick Reference Guide places A2L "Testlets" in the categories of formative, progress monitoring, and diagnostic. These are *assessments for / as learning*, according to the SBE Implementation Plan.

Purveyors of A2L theorize that when schools do a good job of implementing A2L, teachers are more likely to use A2L to inform instruction, and that student achievement should improve as a result. However, anecdotal information suggests that A2L implementation varies across the district and within individual schools. Attendance at A2L training sessions has been poor, with about 25% of all sessions (over 30) being cancelled in 2007-08 due to lack of enrollment. Satisfaction with and commitment to A2L appear to vary widely.

## Research Questions

This study seeks primarily to address whether there is an association between teachers' use of A2L results and student achievement. The study also seeks to describe the nature, frequency and quality of A2L use by APS teachers. Research questions are as follows:

1. Do teachers who use A2L results have higher levels of student achievement than teachers who do not use A2L results?
2. What are the nature, frequency and quality of A2L use by APS teachers? What are the main barriers to greater use?
3. Do teachers who use A2L results perceive that, as a result of A2L, their instruction has improved?

## Method

### Instrument and Sample

RDA used a ten-item survey to collect data from teachers on A2L use and perceived benefits and barriers (see Appendix A). During district-wide professional development workshops in September 2007, RDA staff distributed surveys to teachers who taught 4<sup>th</sup> through 8<sup>th</sup> grade math and reading during 2006-07. Nearly all teachers had the opportunity to complete a survey. A total of 777 4<sup>th</sup> through 8<sup>th</sup> grade teachers completed usable surveys, for an estimated response rate of 44%.<sup>1</sup> (See Appendix B for a frequency report.)

Of the 777 responding teachers, 403 (286 elementary, 26 middle school reading, and 91 middle school math) were linked with students for outcome analysis. The outcome sample consists of students in grades 4 through 8 during the 2006-2007 school year.

### Measures

Outcome measures are individual students' raw A2L reading and math scores from spring 2007. Raw scores range from 0 to 100 and represent the percentage of correct items. Tests vary somewhat in length depending on grade; however, math tests contain about 47 questions while reading tests contain about 42 questions. Questions are not weighted for difficulty.

We measure our independent variable of interest, A2L use, in four ways:

*Fall to Winter Comparison.* This has dichotomous values (Yes, No) and measures whether teachers compared their class's winter 2007 A2L results to the class's fall 2006 A2L results.

*Report Type.* Teachers reported whether or not they obtained each of seven A2L report documents. We used expert opinion to determine the relative helpfulness of each report for tailoring instruction. Reports worth one point include: Class List and Test History by Class; reports worth two points include: Skill Gap Analysis, Skill Proficiency by Class, and the actual test; reports worth three points include: Item Analysis and Skill Proficiency by Student. Teachers were scored on a 0-3 scale according to the highest level of report they used. If a teacher did not use any report, he or she was assigned a score of 0.

*Frequency of Use.* Teachers were scored on a 0-3 scale according to the total number of testing windows they accessed A2L reports (fall, winter, and spring). If teachers never accessed any report, he or she was assigned a score of 0.

---

<sup>1</sup> The survey response rate of 44% is a rough but conservative approximation, because the denominator of 1,758 is an estimate rather than a precise count of all teachers of 4<sup>th</sup> – 8<sup>th</sup> grade reading and math. SchoolNet/AIMS provides the current total number of 4<sup>th</sup> and 5<sup>th</sup> grade teachers but does not identify which ones provide direct instruction in reading or math. We use the total number of 4<sup>th</sup>-5<sup>th</sup> grade teachers, which likely deflates the response rate. SchoolNet/AIMS provides an unduplicated count of middle school math teachers, but reading teachers are identified by an assortment of subject area names, including English, Language Arts, Literature, and LA. We calculated the proportion of overlaps for a sample of these teachers and then estimated an unduplicated count of 6<sup>th</sup>, 7<sup>th</sup> and 8<sup>th</sup> grade reading teachers; our estimate for middle school reading teachers may be high or low.

*Use Activities.* Teachers reported whether or not they used A2L results for a series of activities. We used expert opinion to determine the relative helpfulness of each activity in tailoring instruction. Activities worth one point include: conference with parents and/or teachers, determine AIP eligibility, and send results home. Activities worth two points include: identify gaps in knowledge and understanding, group students, and reflect on effectiveness of teaching strategies. Only one activity is worth three points: plan what skills and processes to teach. Teachers were scored by adding the scores of 2- and 3-point activities. If a teacher reported only 1-point activities or none at all, he or she was assigned a score of 0.

From these four measures of use, we also developed a single composite measure, with dichotomous values, identifying very high users. This measure indicates teachers who compared their class's winter A2L scores to its fall A2L scores and who scored a '3' or higher for report type, frequency of use, and use activities.

A number of student-level control variables are included in our analysis: fall 2006 A2L scores, gender, and status as a special education student, gifted student, English language learner, and/or participant in the Free/Reduced Price Lunch program.

## **Analyses**

Individual students are the unit of analysis for answering Research Question #1. Ordinary Least Squares (OLS) regression analyses are used. In order to account for the hierarchical nature of our data, we cluster students by teacher in all OLS analyses.<sup>2</sup>

Descriptive survey results, including those related to Research Questions #2 and #3, are derived from the full sample of 777 responding 4<sup>th</sup> to 8<sup>th</sup> grade teachers of math and reading.

## **Results**

### **Research Question #1: Do teachers who use A2L results have higher levels of student achievement than teachers who do not use A2L results?**

We use our four measures of A2L use to predict spring 2007 A2L scores while controlling for fall 2007 A2L performance and several demographic variables known to impact achievement. As seen in Table 1, fall to winter comparisons appear to be important to achievement outcomes at least in middle school. Students with teachers who compared A2L results over time earned on average about four points higher on their spring A2L math assessment and 17 points higher on their spring A2L reading assessment. However, fall to winter comparisons do not seem to be important to elementary school students, though the lack of much variance for this measure by elementary school teachers may be concealing its true effect.

In one of the four tests, report type impacts A2L scores in a statistically significant and positive way – elementary reading scores. Similarly, frequency of use is statistically significantly positive for elementary reading scores, but statistically significantly *negative* for middle school reading scores. Of the four measures of A2L use, both report type and frequency of use seem to have

---

<sup>2</sup> Clustering students is a method of avoiding the inflation of the significance of teacher-level variables that can occur. Clustering determines and acknowledges that some of that variance between teachers is due to unmeasured factors. Clustering does not change coefficients, but rather increases standard errors.

lower face validity than fall to winter comparisons and use activities. Additional research will be necessary to learn whether these findings are statistical artifacts or whether they do indeed impact individual students' test scores.

Finally, we find no evidence that use activities impact achievement. This result is unexpected, for higher scores on this measure are meant to capture activities that APS experts suppose to be most helpful for adjusting instruction, such as identifying gaps in knowledge and planning and grouping students for instruction.

Table 1. Impact of A2L Use on Student Spring 2007 A2L Scores

	Elementary Math	Elementary Reading	Mid School Math	Mid School Reading
<i>n</i> (teachers)	250	249	72	17
<i>n</i> (students)	4,822	4,759	3,138	945
<i>r</i> <sup>2</sup>	.48	.57	.54	.61
<i>A2L Use</i>				
Fall to Winter Comparison	.38 (1.52)	-.17 (1.18)	4.26 <sup>†</sup> (1.82)	17.30* (1.63)
Report Type	.67 (.54)	.82* (.39)	.65 (.91)	-.73 (.52)
Frequency of Use	1.24 (.87)	1.54* (.73)	-.16 (1.06)	-4.79* (1.09)
Use Activities	-.03 (.19)	-.24 (.14)	-.14 (.35)	-.08 (.24)
<i>Control Variables</i>				
Fall A2L score	.66* (.02)	.66* (.01)	.71* (.02)	.60* (.02)
English language learner	-.99* (.23)	-1.87* (.23)	-1.19* (.42)	-2.57* (.49)
Special Education	-6.99* (1.02)	-5.77* (.78)	-3.24* (1.06)	-5.17* (2.33)
Gifted	5.43* (.64)	4.57* (.62)	6.56* (1.16)	5.66* (1.18)
Free/Reduced Price Lunch	-4.50* (.66)	-3.84* (.57)	-2.28* (.66)	-5.06* (.60)
Female	-.09 (.35)	1.07* (.34)	.23 (.44)	2.26* (.85)

<sup>†</sup>p<.01, \* p<.05

Overall, these results suggest that the impact of using A2L results is limited. None of our measures of A2L use has consistent effects on student achievement scores. These measures, of course, are limited. For example, they do not address the quality or accuracy of using A2L information.

**Research Question #2: What are the nature, frequency and quality of A2L use by APS teachers? What are the main barriers to greater use?**

*Obtaining Assessment Results – Frequency of Use and Report Types*

Survey results suggest that most teachers (69%,  $n=680$ ) obtain A2L results for every test window (fall, winter, spring). Teachers are most likely to obtain the Class List report, which lists overall test scores by student. Almost half (44%) obtain one or more detailed reports, such as the Skill Proficiency by Student, Skill Proficiency by Class, Item Analysis and Skill Gap reports. Table 2 shows the types of A2L reports and the percentages of teachers who report obtaining each one during 2006-07.

Table 2. Percent Teachers by Type of A2L Report Obtained

<b>A2L Report</b>	<b>Percent Teachers (<math>n = 777</math>)</b>
Class List	59%
Skill Proficiency by Student	29
Item Analysis	22
Skill Proficiency by Class	19
Skill Gap Analysis	15
Actual Test	11
Test History by Class	7
No report obtained	24

*Teachers' A2L Use Activities*

Survey results suggest that most teachers use A2L results in some way, but few use A2L to help modify classroom practices. At the most basic level, many teachers say they compare their class's fall and winter A2L results (61%,  $n=755$ ). Most discuss results with their grade level or department at least one time during the year (77%,  $n=755$ ), and almost half (47%) discuss results three or more times.

The most common application of A2L results is to determine AIP eligibility, cited by 57% of teachers.<sup>3</sup> About half of teachers report using A2L results to inform their instructional practice in at least one way: to identify gaps in student knowledge and understanding (39%); to plan what skills and processes to teach (36%), to reflect on the effectiveness of teaching strategies (22%), and/or to group students (14%). Less than half discuss A2L results with parents and students or send results home.

---

<sup>3</sup> The AIP website determines AIP eligibility automatically using A2L results.

Table 3. Percent Teachers by Type of A2L Use

Type of A2L Use	Use Level	Percent of A2L users (n = 777)
Determine AIP eligibility	Low	57%
Inform instructional practice in at least one way:		52
• Identify gaps in knowledge & understanding	Medium	39
• Plan what skills & processes to teach	High	36
• Reflect on effectiveness of teaching strategies	Medium	22
• Group students	Medium	14
Conference with parents/students or sent results home	Low	46

Table 4 reveals the relative infrequency with which teachers use A2L assessment results to adjust classroom practices, perhaps the real test of A2L use. To “earn” 7 points, a teacher had to report using A2L reports to plan what to teach, plus at least two of these other purposes: to group students, to identify gaps in knowledge/understanding, and/or to reflect on the effectiveness of teaching practices. Few teachers meet or exceed this bar – just one quarter (25%) of elementary school teachers, 14% of middle school math teachers, and only 8% of middle school reading teachers.

#### *Diagnostic Testlets*

Only nine percent of responding teachers report administering A2L diagnostic testlets. Testlets are designed to explore students’ comprehension and skills in greater depth, and to monitor progress in between A2L survey tests. Testlet results may be accessed within the instructional segment being taught, thus allowing teachers to modify or augment instruction before the unit is over. This is the kind of assessment that research has linked with improved student achievement. Expanding the use of A2L diagnostic testlets may therefore be a strategy the district wishes to explore.

*Use by School Level*

Table 4 compares the elementary and middle school teacher groups for each measure of A2L use. There are two important trends to notice in this table. Most obviously, elementary teachers are much more likely to use A2L results than middle school teachers. And among middle school teachers only, math teachers are much more likely to use results than reading teachers. For each of the four measures of use, over two-thirds of all elementary teachers reported high use. On the other extreme, under half of middle school reading teachers reported high use on three of the four measures: report type, frequency of use, and use activities. About 16% of elementary teachers are “very high users” according to the composite measure, while only 10% of middle school math teachers are very high users. No middle school reading teachers in our sample qualify as high users.

Secondly, some measures of A2L use are more common than others. Overall, three out of four teachers access reports for all three testing windows and over 80% compare fall and winter reports. More meaningful measures of A2L use are less widespread. Fewer than half of teachers obtain either detailed 3-point report, the Item Analysis or Skill Proficiency by Student. And only about one in five teachers use A2L in multiple ways deemed most helpful for tailoring instruction.

Table 4. Percent Teachers by Measure of A2L Use<sup>4</sup>

	<b>Fall to Winter Comparison:</b> % teachers indicating 'yes'	<b>Report Type :</b> % teachers that obtained a 3-point report	<b>Frequency of Use:</b> % teachers that accessed reports for all three testing windows	<b>Use Activities:</b> % teachers whose sum of activities is worth at 7 points	<b>Composite Measure :</b> % teachers high on all four measures of use
Elem Teachers	77% (n=255)	57% (n=256)	85% (n=256)	25% (n=288)	16% (n=284)
MS Math Teachers	66% (n=73)	27% (n=89)	74% (n=55)	14% (n=91)	10% (n=91)
MS Reading Teachers	47% (n=17)	20% (n=25)	44% (n=18)	8% (n=26)	0% (n=26)

<sup>4</sup> Table 4 reflects the sub-sample of 403 teachers for whom course/subject data were available.

### *Barriers to Using A2L Results*

Survey respondents were asked to name the main barrier to using A2L results more than they did. As shown in Table 5, technical difficulties is the most commonly cited top barrier, followed by difficulty in accessing reports, not having enough time, and not knowing how to access reports. Medium and high users of A2L results are most likely to cite technical difficulties. Non-users are most likely to cite lack of knowledge. Only small fractions of respondents name concerns about the validity, interpretability or usefulness of A2L results. The smallest fraction blames school priority setting.

Table 5. Percent Teachers by Cited Top Barrier to Greater A2L Use

<b>Top Barrier to A2L Use</b>	<b>Percent Teachers (n = 580)<sup>5</sup></b>
Technical difficulties	21%
Reports difficult to access	15
Not enough time	13
Didn't know how to access A2L reports	12
Other barrier	11
Didn't think results were valid	6
Reports difficult to interpret	3
Didn't think results were useful	2
A2L was not a priority at my school	2

### **Research Question #3: Do teachers who use A2L results perceive that, as a result of A2L, their instruction has improved?**

Almost all teachers who use A2L results at the highest level agree that A2L improves their instructional practice in at least one way (96%,  $n = 76$ ). Most agree that it helps them tailor instruction and address a wide range of learning needs. Seventy percent say it helps them improve their overall instructional effectiveness.

Table 6. Percent of High A2L Users by Reported Instructional Benefit

<b>Instructional Benefit</b>	<b>Percent of High Users (n = 76)</b>
A2L helped me do a better job of addressing a wide range of learning needs.	78%
A2L helped me do a better job of tailoring instruction to individual learning needs.	80
A2L helped me improve my effectiveness as a teacher.	70

<sup>5</sup> This sample includes only respondents that specified one top barrier. Those who named more than one top barrier were excluded from analysis.

## Limitations

### Defining Level of Use

Obtaining an accurate estimate of the level at which each responding teacher used A2L results in 2006-07 is essential for answering the primary research question. The survey contains six questions addressing different aspects of use, including: which reports were obtained, whether teachers compared fall and winter results, how often they discussed results with their grade level or department, how they used results, and whether they administered diagnostic tests. The following measurement limitations are important to note:

- The survey asked teachers to identify the ways they used A2L results, but it did not collect information on the depth, frequency or breadth of use. Consequently, a teacher who used A2L results to plan instruction once during the year generated the same A2L use score as a teacher who used A2L results to plan instruction on a weekly basis.
- Self-reported data are inherently limited by biases related to recall, judgment, social desirability and human error. Validity checks discovered many conflicting responses within and between items. For example, some teachers marked one or more ways they used A2L results but also reported not using A2L results. RDA eliminated from analysis all conflicting responses that could be identified, but it is possible that other validity problems remained.

Our ability to speak with confidence about results and issue conclusions is constrained by the following sampling and study design limitations:

- High/proficient users of A2L are likely to be high/proficient users of other assessment and instructional processes. This study did not collect data on other aspects of instruction. Therefore, it is impossible to know how much A2L use contributed, if at all, to a good teacher's overall effectiveness.
- The sample of middle school reading teachers is inexplicably small ( $n=26$ ), and represents only about six percent of middle school reading teachers. In addition, we have school locations for only half of all survey respondents (teachers who provided accurate employee numbers). These limitations constrain our ability to estimate the representativeness and generalizability of survey results.

## Discussion

This study validates concerns that A2L is underused by a majority of APS teachers. About half of teachers report using A2L results to inform their instruction, but only about 16% of elementary teachers, 10% of middle school math teachers and no middle school reading teachers qualify as “very high users” of A2L.

An encouraging finding is that 96% of high-level A2L users perceive that A2L improves their instructional practice. Extending A2L's potential benefits to other teachers may require targeted professional development and technical assistance. Sixty percent of low A2L users say the main barrier to using A2L is that reports are difficult to access or that they do not know how to access A2L reports. Poor attendance at A2L training sessions could be to blame. To increase the number of teachers attending A2L workshops, APS may need to modify its training approach,

consider incentives, and/or integrate A2L into more instructionally oriented professional development workshops.

The focus of this study is whether A2L can generate improvements in student achievement. The literature on formative assessment provides compelling evidence that teachers who integrate assessment into their everyday instructional practice do, indeed, help their students achieve significant learning gains. Black and Wiliam (1998) report that formative classroom assessment practices produce average effect sizes between 0.4 and 0.7<sup>6</sup> and that the largest gains occur for the lowest achievers.<sup>7</sup> Subsequent studies have concluded that formative assessment is more cost-effective in raising student achievement than class-size reduction, increases in teacher content knowledge, increased per-pupil expenditures, increased accountability, and other popular educational initiatives.<sup>8</sup>

W. James Popham encourages educators to distinguish carefully between formative classroom assessments and interim benchmark assessments. Formative classroom assessments yield results during a class period or in the midst of a multiweek instructional unit. Tests that provide results for topic X when the teacher has already moved onto topic Z cannot be regarded as formative assessment.<sup>9</sup> A2L diagnostic testlets may be accessed and used within the instructional segment being taught. This component of A2L, therefore, may be a powerful tool for advancing student achievement -- if teachers use testlets to actively engage students in assessing progress toward learning targets, and if they provide descriptive feedback and guidance about exactly what steps students can take to improve.<sup>10</sup>

A2L survey tests fit into the category of interim benchmark assessments, which teachers may use for longer-term formative purposes. Results from the present study suggest that comparing fall and winter A2L scores may help teachers and students, at least at the middle school level. However, the formative use of benchmark tests like A2L alone may not be sufficient to impact student achievement. This finding mirrors W. James Popham's conclusion that interim or benchmark tests have no evidence of short term benefit.

---

<sup>6</sup> At the low end, this means formative assessment raised the average intervention student's achievement to the level of the top 35% of non-intervention students.

<sup>7</sup> Black, P., Wiliam, D. (1998). Inside the Black Box: Raising Standards Through Classroom Assessment. *Phi Delta Kappan*, 80 (2). Retrieved April 22, 2008 from <http://www.pdkintl.org/kappan/kbla9810.htm>.

<sup>8</sup> Wiliam, D. (2006, July) *Does Assessment Hinder Learning?*. Speech delivered at the ETS Breakfast Salon. Retrieved April 21, 2008 from [http://www.uk.etseurope.org/fileadmin/free\\_resources/UK%20website/UK\\_Current/Dylan\\_Wiliam\\_speech.doc](http://www.uk.etseurope.org/fileadmin/free_resources/UK%20website/UK_Current/Dylan_Wiliam_speech.doc)

<sup>9</sup> Popham, W. J. (2006). Phony Formative Assessments: Buyer Beware! *Educational Leadership*, 64(3), 86-87.

<sup>10</sup> Chappuis, J. (2005). Helping Students Understand Assessment. *Educational Leadership*, 63(3), 39-43. Retrieved April 22, 2008 from <http://www.iowa.gov/educate/content/view/1124/1490>

# Appendix A

## Teacher Survey of A2L Use

Survey of Assess2Learn Use

**PLEASE COMPLETE FOR IMMEDIATE COLLECTION**

**For all 4th - 8th Grade Teachers  
of Reading and Math in 2006-07**

**Purpose:** This survey is designed to measure whether and how APS teachers use the Assess2Learn assessment. Survey results will help RDA understand how well A2L is working as well as how it can be made more effective.

**Honest responses will help improve A2L.** For many reasons, you may use A2L regularly, very little or not at all. For example, your school's computers may or may not have problems. You may or may not have received training in how to use A2L results. You may or may not use other assessments to drive your instruction. For you and many other teachers, "no," "none," or "never" may be appropriate answers to survey questions.

Teacher ID Number					
0	0	0	0	0	0
1	1	1	1	1	1
2	2	2	2	2	2
3	3	3	3	3	3
4	4	4	4	4	4
5	5	5	5	5	5
6	6	6	6	6	6
7	7	7	7	7	7
8	8	8	8	8	8
9	9	9	9	9	9

**PLEASE USE PENCIL OR BLUE OR BLACK PEN ONLY TO FILL OUT SURVEY.**

**All your answers will be confidential.** No one will be able to trace your responses to you. RDA researchers will use teacher ID numbers ONLY to identify student groups. **All teacher and student ID numbers will be eliminated immediately** after data groups are formed.

**Please answer the following questions about the 2006-2007 school year.**

	Fall Results	Winter Results	Spring Results	None
1. Which A2L results did you obtain last year (2006-07), if any? <b>Mark all that apply.</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

2. Which A2L reports did you obtain last year? <b>Mark all that apply.</b> <ul style="list-style-type: none"> <li>a. <input type="radio"/> Class List (Test scores by students)</li> <li>b. <input type="radio"/> Item Analysis (Class responses for each item, organized by skill)</li> <li>c. <input type="radio"/> Skill Gap Analysis (For each skill, a list of students with percent items answered correctly)</li> <li>d. <input type="radio"/> Skill Proficiency by Student (For one student, by skill, percent items answered correctly)</li> <li>e. <input type="radio"/> Skill Proficiency by Class (For each skill, percent items answered correctly by class as a whole)</li> <li>f. <input type="radio"/> Test History by Class (Number tests taken and average, minimum and maximum scores for each student)</li> <li>g. <input type="radio"/> Actual Test</li> <li>h. <input type="radio"/> None</li> </ul>
--

	Yes	No
3. Did you have students take optional A2L diagnostic tests in addition to required fall, winter & spring tests?	<input type="radio"/>	<input type="radio"/>

RDA/dh - 8/07

**TURN OVER**

## Appendix A Teacher Survey of A2L Use

4. If you used A2L results last year (2006-07), how did you use them? **Mark all that apply**

- a.  Identify gaps in knowledge and understanding
- b.  Plan what skills and processes to teach
- c.  Group students
- d.  Reflect on effectiveness of teaching strategies
- e.  Conference with parents and/or students
- f.  Determine AIP eligibility
- g.  Sent results home
- h.  Other
- i.  I did not use A2L results last year

	Yes	No
5. Did you compare your own class's Winter A2L results to the class's Fall A2L results?	<input type="radio"/>	<input type="radio"/>

	Never	Once	Twice	Three or More Times
6. How often did you discuss A2L results with your grade level or department last year?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

7. What was the main barrier to using A2L results more than you did last year? **Mark the ONE** most important barrier.

- a.  Technical difficulties
- b.  Reports difficult to interpret
- c.  Reports difficult to access
- d.  Didn't know how to access A2L reports
- e.  Didn't think results were useful
- f.  Didn't think results were valid
- g.  Not enough time
- h.  A2L was not a priority at my school
- i.  Other barrier
- j.  I used A2L results to their maximum capacity

	Strongly Disagree	Disagree	Agree	Strongly Agree
8. A2L helped me do a better job of addressing a wide range of learning needs.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9. A2L helped me do a better job of tailoring instruction to individual learning needs.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10. A2L help me improve my effectiveness as a teacher.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**DELIVER TO SURVEY COLLECTION BOX AT FRONT OF ROOM**

**THANK YOU!**

Contact Debra Heath at 848-8724 if you have any questions.

## Appendix B

	Elementary school teachers	Middle school math teachers	Middle school reading teachers
<i>n</i>	284	91	26
1. Which A2L results did you obtain last year?			
Fall	84%	67%	35%
Winter	81%	64%	31%
Spring	88%	81%	50%
None	6%	14%	42%
2. Which A2L results did you obtain last year?			
Class List	76%	71%	50%
Item Analysis	34%	17%	0%
Skill Gap Analysis	25%	13%	4%
Skill Proficiency by Student	44%	19%	19%
Skill Proficiency by Class	34%	7%	12%
Test History by Class	11%	9%	15%
Actual Test	11%	9%	4%
None	11%	16%	38%
3. Did you have students take optional A2L diagnostic tests?			
Yes	9%	6%	12%
4. If you used A2L results last year, how did you use them?			
Identify gaps in knowledge and understanding	53%	42%	19%
Plan what skills and processes to teach	50%	39%	19%
Group students	18%	11%	15%
Reflect on effectiveness of teaching strategies	31%	24%	12%
Conference with parents and/or students	55%	43%	23%
Determine AIP eligibility	78%	65%	27%
Sent results home	29%	24%	4%
Other	4%	6%	4%
Did not use A2L results last year	8%	12%	35%
5. Did you compare your class's winter A2L results to the class's fall A2L results?			
Yes	77%	66%	47%

	Elementary school teachers	Middle school math teachers	Middle school reading teachers
6. How often did you discuss A2L results with your grade level or department last year?			
Never	11%	14%	32%
Once	11%	9%	9%
Twice	22%	22%	23%
Three or more times	56%	55%	36%
7. What was the main barrier to using A2L results more than you did last year?			
Technical difficulties	33%	36%	19%
Reports difficult to interpret	4%	7%	12%
Reports difficult to access	33%	19%	19%
Didn't know how to access A2L reports	15%	19%	19%
Didn't think results were useful	6%	4%	8%
Didn't think results were valid	15%	4%	8%
Note enough time	24%	24%	19%
A2L was not a priority at my school	1%	1%	4%
Other barrier	9%	7%	4%
I used A2L results to their maximum capacity	5%	3%	0%
8. A2L helped me do a better job of addressing a wide range of learning needs.			
Strongly disagree	14%	15%	14%
Disagree	37%	34%	50%
Agree	45%	46%	32%
Strongly agree	4%	5%	4%
9. A2L helped me do a better job of tailoring instruction to individual learning needs.			
Strongly disagree	13%	13%	14%
Disagree	38%	40%	50%
Agree	44%	44%	32%
Strongly agree	5%	4%	4%
10. A2L helped me improve my effectiveness as a teachers.			
Strongly disagree	15%	20%	23%
Disagree	40%	36%	36%
Agree	41%	40%	32%
Strongly agree	4%	4%	9%