

# **Reading Verification Improvements in Scientific Learning Reading Assistant™ Expanded Edition**

**Scientific Learning: Research Reports, 13(13):1-18**

<b>Executive Summary .....</b>	<b>2</b>
<b>Introduction.....</b>	<b>4</b>
<b>New Acoustic Models for Regional Dialects (Southern U.S.).....</b>	<b>4</b>
<b>Changes to Intervention Timing.....</b>	<b>8</b>
<b>Words Correct Per Minute Calculation .....</b>	<b>12</b>
<b>Default Glue Word List Changes.....</b>	<b>14</b>
<b>Conclusion .....</b>	<b>18</b>

## **Executive Summary**

Scientific Learning Reading Assistant™ Expanded Edition has incorporated a number of improvements in its ‘Reading Verification’ technology. This technology is what enables the software to perform as a personal tutor for guided oral reading practice. It has been revised for Reading Assistant Expanded Edition, based on best practices in reading instruction as well as empirical data from hundreds of readers. The improvements allow the software to tutor a wider range of readers, provide a better user experience for students, and provide more useful feedback for teachers.

For oral reading practice to be constructive and motivating, readers need support when they are struggling or making errors. However, they also need opportunities to practice sustained reading without too many interruptions, so when they are reading well they shouldn’t receive too many interruptions. To meet both of these needs, a tutor must be an accurate listener. To ensure greater accuracy for more readers, new speech models have been developed for southern regional dialects. Based on speech data collected from hundreds of children and adults in the region, these new models make Reading Assistant Expanded Edition a more accurate listener.

Timely intervention is critical for a tutor supporting oral reading practice. Interventions that occur too soon or too late can disrupt fluency, reduce opportunities for learning, and frustrate the reader. Data collected from struggling readers was used to refine the timing of interventions in Reading Assistant Expanded Edition. The new design will help improve the accuracy and timing of the interventions for each reader. A special set of rules is applied to less fluent readers, so that they receive extra time for decoding. Reading Assistant Expanded Edition will also accommodate less fluent readers by providing more support for very high frequency “glue words.”

To aid in progress monitoring, Reading Assistant Expanded Edition reports on several aspects of reading performance including comprehension skills and strategies, problem words, and reading rate. The method for calculating reading rate scores has been revised in Reading Assistant™ Expanded Edition, for closer alignment with scores from traditional fluency measures. This improved alignment makes the revised Words Correct Per Minute (WCPM) score easier for teachers to interpret.

Compared to earlier versions of the software, Reading Assistant Expanded Edition is better at listening accurately, providing timely interventions, and accommodating less fluent readers. It also provides reading rate scores that are more closely aligned with conventional fluency measures, so that teachers can better monitor student progress and intervene appropriately. All of these improvements help Reading Assistant Expanded Edition provide an optimal learning environment for guided oral reading practice.

## **Introduction**

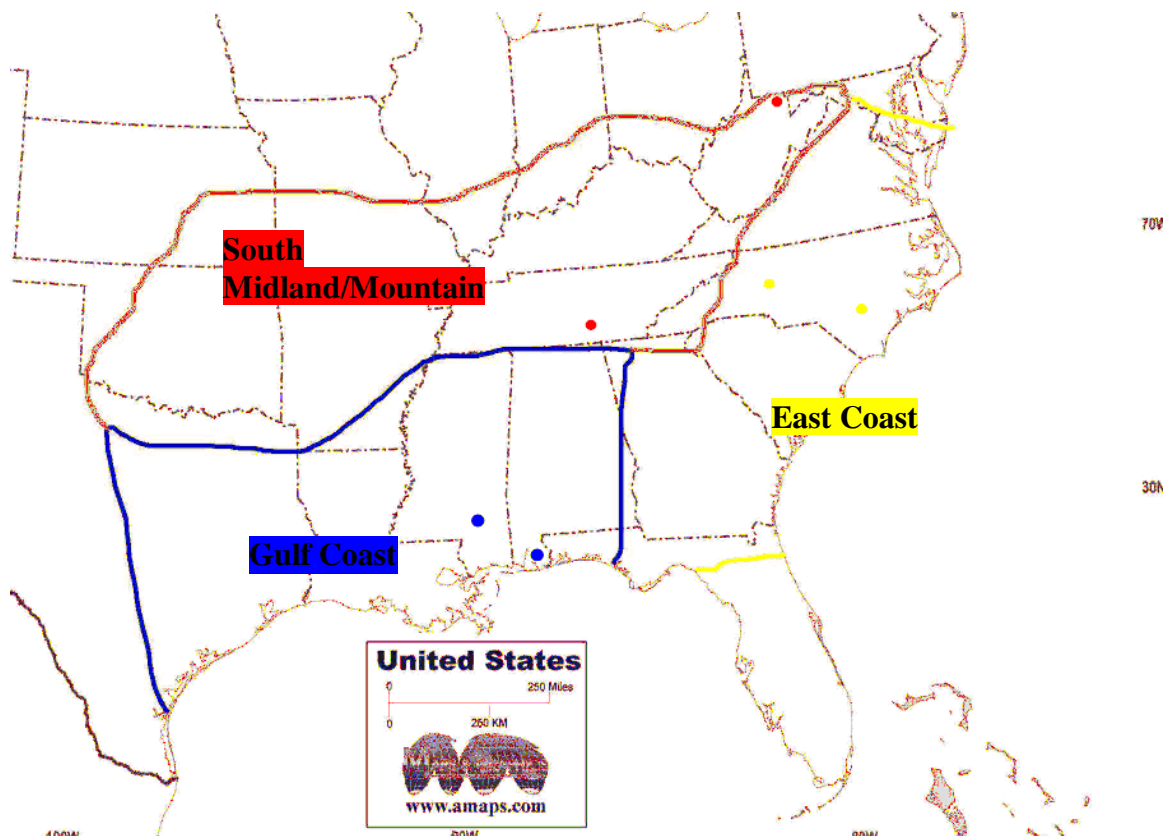
Reading Assistant™ Expanded Edition (RAEE) incorporates several significant new improvements to its ‘Reading Verification’ technology. This technology is what enables Reading Assistant to act as a helpful listener in ‘Record My Reading’ mode and provide feedback to students and teachers. The components that provide this capability include the speech recognition engine, the user interface for ‘Record My Reading’ mode, and the logic behind the fluency measurements and ‘problem word’ lists. Improvements have been made in all three of these areas, with the goal of enhancing the user’s guided oral reading experience, while supporting the range and structure of the expanded content sets. This report describes each of these improvements in detail and provides results quantifying the improvements, where appropriate.

### **New Acoustic Models for Regional Dialects (Southern U.S.)**

The largest and most noticeable impact on the usability and performance of the Reading Verification technology comes from the addition of new acoustic models focused on regional dialects of the Southern United States. Acoustic models represent the sounds and sound sequences of spoken language and are an essential and critical part of the speech recognition software. During the initial voice customization process in Reading Assistant Expanded Edition, the best acoustic model is chosen from the available model set. If none of the available models match the user very well, speech recognition performance will be suboptimal. The models available in earlier versions of Reading Assistant did not represent the pronunciations of the Southern region very well, and some performance problems were noted for users in this region.

To provide the necessary data for building new models, audio data from 786 subjects was collected at six locations in the Southern region. The data gathering effort used a configuration of the Reading Assistant Version 4 software customized for data collection. After setting aside data that did not meet dialect or audio quality criteria, this collection resulted in over 110 hours of audio data from 685 subjects, which was used to create new

adult and child acoustic models. An important goal of the data collection was to get a distribution of subjects across age, gender, and dialect region within the South. Age, gender and dialect region all influence the speech characteristics of the subject and it is important to get this variability represented in the acoustic models. For the purposes of the collection, the Southern U.S. was divided into three dialect sub-regions, with two data collection sites in each region as shown in Figure 1.



*Figure 1. Dialect Sub-Regions of the Southern U.S. Dots represent data collection locations. (Original map source: [http://www.amaps.com/mapstoprint/OUTLINE%20MAPS/free\\_map\\_of\\_usa.htm](http://www.amaps.com/mapstoprint/OUTLINE%20MAPS/free_map_of_usa.htm))*

The breakdown of subjects by age range, gender, and dialect region for the 685 subjects used to develop the models is shown in Tables 1, 2 and 3.

<i>Age</i>	<i>6-8 years old</i>	<i>9-11 years old</i>	<i>12-14 years old</i>	<i>15-17 years old</i>	<i>Adult</i>
Number of subjects	131	152	165	113	124

*Table 1. Number of Subjects by Age Range*

<i>Gender</i>	<i>Female</i>	<i>Male</i>
Number of subjects	356	329

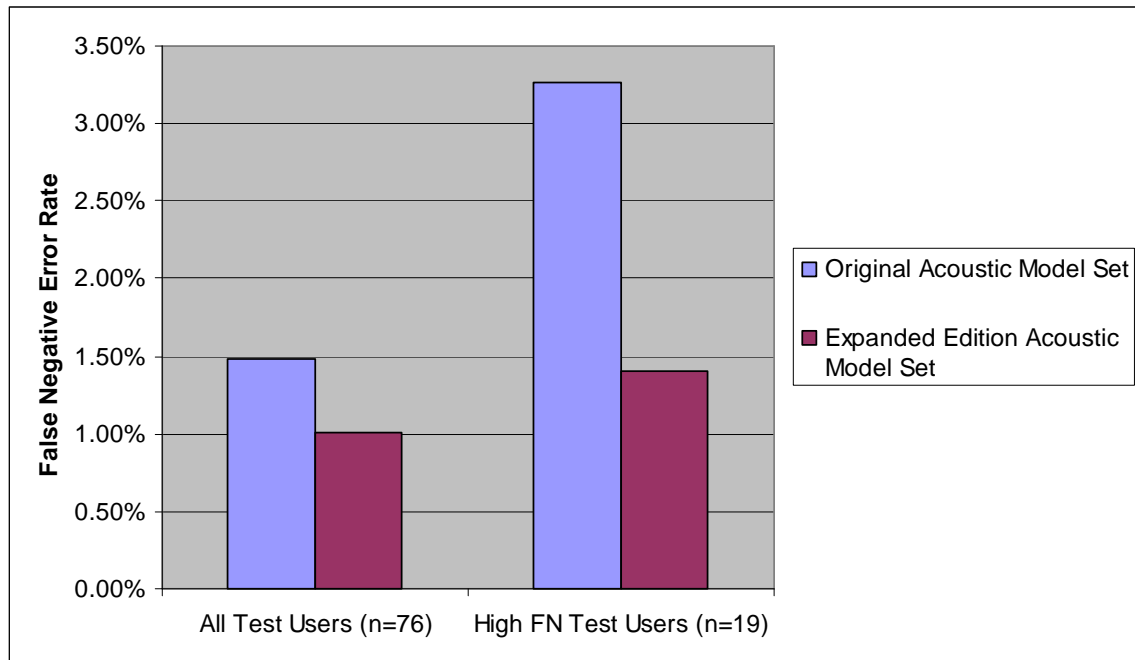
*Table 2. Number of Subjects by Gender*

<i>Age</i>	<i>East Coast</i>	<i>Gulf Coast</i>	<i>South Midland/Mountain</i>
Number of subjects	156	265	264

*Table 3. Number of Subjects by Southern Sub-region*

The goal of creating these new models was to decrease the false negative error rate for Southern speakers without adversely impacting other error metrics. The false negative rate represents how frequently a user is stopped by the software on a word they read correctly. In general, we aim for an average false negative rate of about 1% (1 in 100 words). In our experience, if a user is stopped on a correctly read word more than 2% of the time (1 in 50 words, or perhaps about twice on a page spread of text), it is more likely that the user will be frustrated and fluency may be disrupted.

For the new models, the goal was to decrease the false negative rate by at least 30% overall for Southern speakers, compared to performance on the existing acoustic model set. The performance for users with high error rates on the existing acoustic model set was also analyzed. Performance was analyzed using a set of 76 speakers from the data collection whose data was not used to create or adjust parameters for the acoustic models but was instead ‘held out’ as test data.



*Figure 2. False Negative Error Rate Reduction Including the Southern Acoustic Models/ High FN Test Users are users with greater than 2% false negative error rate when using the original (Reading Assistant Version 4) acoustic model set*

Figure 2 shows the results in terms of false negative (FN) error rate for the original model set compared to the new model set. For these results, the voice customization enrollment data was used to select the best available model for each user, and then that model was used for the remaining testing on ‘Record My Reading’ audio recordings. To test Reading Assistant™ Version 4 performance, the model set used in voice customization was the original model set consisting of child, adult male, and adult female models. To test Reading Assistant Expanded Edition performance, the model set used was the original model set plus the new Southern child and Southern adult models (five models total in the set). Thus the automated testing design emulates actual real-life performance as closely as possible.

These results indicate that we met the 30% reduction goal, bringing the false negative rate down to approximately 1%, on average. The results are also broken out separately for users who are ‘outliers’ as far as performance with the original models, defined as users having a false negative rate of greater than 2% with the original models. For these 19 users the results are much more dramatic: the false negative error rate is reduced by more than a factor of two. The number of speakers in the total set with an error rate of greater than 2% is reduced from 19 with the original model set, to 6 with the new (RAEE) model set.

Another important error metric is the false positive rate. The false positive rate represents how often we allow the user to proceed when an error has been made. Since false positive and false negative rates trade off against one another in a system, it is important to measure both rates to understand if we have really improved overall performance. After incorporation of the new acoustic models, the false positive rate is slightly improved, with a reduction of approximately 1%. In other words, we have reduced the false negative rate significantly while keeping the false positive rate essentially the same, resulting in a significant improvement to overall performance for the Reading Verification task.

### **Changes to Intervention Timing**

Reading Assistant™ Expanded Edition contains more lower-level content and content for below grade level readers than did earlier versions of Reading Assistant. As part of this shift in focus, adjustments were made to the timing parameters for interventions. The basic motivation for these changes was the pedagogical principle that readers at the lower levels should be given more time to work on and struggle with a word, relative to more advanced levels. To determine the best timing for different readers, extensive analyses of empirical data were conducted. Informed by both theory and data, the intervention timing parameters were made dependent on text reading level, and were made longer for lower reading levels relative to the settings for earlier versions of Reading Assistant.



Reading Assistant™ will give an intervention (visual and/or audio feedback on a word) when it detects that the child has not read a word or has read it incorrectly. There are two levels of intervention given during “Record My Reading” mode: if we have not heard the user read a word and a specified amount of time has elapsed since we heard them read the previous text word, the word is highlighted (no audio). If we still don’t hear the word after a second specified amount of time has elapsed, we maintain the highlight and play an audio prompt for that word.

Two timing parameters were adjusted: the time interval allowed after hearing the previous word before a highlight intervention will be given, and the additional time interval allowed after the highlight intervention before an audio intervention will be given. These settings will be referred to as the ‘highlight intervention interval’, and the ‘audio intervention interval’ in this document. In Reading Assistant Version 4, both the highlight intervention interval and the audio intervention interval were set to 2.0 seconds by default. In other words, if the user just stops reading, a total of 4 seconds will elapse before they will be given the audio prompt.

The data corpus used for automated analysis was the K-2 Data Collection, collected in October 2008 and including 48 students in 2<sup>nd</sup> through 6<sup>th</sup> grade. This data collection used an early version of the RAEE content and focused on below-grade-level readers. The data was collected with a special version of Reading Assistant where the timing parameters for interventions were set to longer values (than the 2 second/ 2 second defaults) so that the intervention timing settings could be analyzed in automated testing. The analysis divided the data into sections by reading level and for each reading level analyzed the relative benefit of increases to these timing parameters, in terms of how often giving additional time resulted in the user coming up with the correct word. On the other side of the equation, lengthening the timing parameters will increase the likelihood of a false positive error (accepting an incorrect attempt, or off-topic or background

speech, as the word). Lengthening the timing parameters will also mean that the user will have to wait longer to get help (an audio prompt) on a word that they are unable to decode by themselves.

The results of the automated analysis, combined with pedagogical recommendations, were used to determine new settings for the lowest text reading levels. The default settings for 2<sup>nd</sup> grade text levels and below were increased to 3.0 seconds for a highlight intervention, and 3.0 seconds between a highlight intervention and audio intervention. The settings of these parameters for 3<sup>rd</sup> grade text levels and higher remain the same (2.0 seconds and 2.0 seconds) as in Reading Assistant™ Version 4.

#### *Intelligent Intervention Logic*

In addition to the elapsed time measurements used to trigger interventions, there is logic in the Reading Verification components that will adjust intervention behavior based on what the user is doing at that moment. For example, if we do not get recognition for a text word (either user error or recognizer error) but the user continues to read, we will intervene faster so that the user can repair to the intervened word before they get too far past the error in the text. If we do not hear the user read a text word, and do not hear them read on past it, but we detect that they are still speaking at the point where we are about to give an intervention, Reading Assistant will ‘defer’ the intervention for a short time period (0.75 seconds). The goal of this logic is to allow a user to finish the word if they are in the middle of it, and avoid ‘interrupting’ if possible. For the most part this intelligent intervention logic remains in place with the new timing settings, with one exception: when using the new 3.0 second/3.0 second defaults for text levels 2-2 and below, an intervention will not be ‘deferred’ even if we detect that the user is still speaking. The reasoning behind this change is that the user has already been given 3 seconds to read the word, and automated analysis suggests that we have reached a point of diminishing returns at that point, so giving further time is not likely to be productive.

### *Implementation Details*

The availability of the K-2 Data Collection and the expansion of automated testing capabilities also provided the opportunity to look at a third timing parameter, the skip interval. If the user still does not read the word after the audio prompt, Reading Assistant™ will wait an additional time interval before ‘skipping’ that word and moving the highlighting on to the next word. In Reading Assistant Version 4, that interval was set to 4.0 seconds. Automated analysis suggested that a value of 3.0 seconds was enough time and that there was no noticeable advantage to giving additional time. Therefore this timing setting was changed to 3.0 seconds for all text reading levels.

Table 4 summarizes the changes made to the intervention timing settings and logic in Reading Assistant Expanded Edition.

<i>Version</i>	<i>Text Reading Level</i>	<i>Highlight Intervention interval</i>	<i>Audio Intervention Interval</i>	<i>Intervention is deferred (if user speaking)</i>	<i>Skip interval</i>
Reading Assistant Version 4	All levels	2.0 sec.	2.0 sec	Yes (0.75 sec.)	4.0 sec.
Reading Assistant Expanded Edition	1-1,1-2, 2-1,2-2	3.0 sec.	3.0 sec	No	3.0 sec.
	3-1 and higher	2.0 sec.	2.0 sec	Yes (0.75 sec.)	3.0 sec.

*Table 4. Old and New Default Intervention Timing Settings for Reading Assistant*

### *Intervention Wait Time Slider*

The default values given in the table above will be appropriate for the majority of users, especially now that the timing parameters are dependent on reading text level. However the ability to adjust these values on a per-user basis is desirable, particularly for users with speech or language disorders, or in other cases where it may be helpful to give a user more time to struggle with words. These values can be adjusted from the default settings via the ‘Intervention Wait Time’ slider which can be accessed from the Reading Assistant

client or Gateway Edition software. Since the default values now vary by text level, the slider in Reading Assistant™ Expanded Edition represents settings on a relative scale ('more time' or 'less time' relative to the current default) rather than giving slider settings in seconds.

### **Words Correct Per Minute Calculation**

The Words Correct Per Minute (WCPM) calculation was adjusted in Reading Assistant Expanded Edition to make it more comparable to a paper assessment of WCPM. Because Reading Assistant is more interactive than a paper assessment, WCPM scores from these two sources should not be expected to correspond exactly. Still, this correspondence is much higher in Reading Assistant Expanded Edition than it was in prior versions of the software.

In a paper assessment, a teacher will intervene (provide the word to the reader) only if the child is completely stuck on a word and is not proceeding in the text. In Reading Assistant, the software will intervene (highlight a word and/or provide an audio prompt) if it detects a significant error, even if the user continues to read. This may happen in the case of user error (e.g. a skipped word), and may also happen if the software misrecognizes a word. In cases where the user has continued to read, this adds time to the total time taken to read a passage, since the user has to back up to the error and repeat from there. This 'time penalty' will tend to result in lower WCPM measurements compared to a paper assessment.

The reading-time measurement used to calculate WCPM has been changed by discounting (removing from the total time taken to read a passage) time corresponding to repeated words, if the repetition immediately follows an intervention. We assume that repetitions following an intervention were 'prompted' or caused by the intervention. Spontaneous repetitions or self-corrections that occur in reading and are not associated

with an intervention will continue to be counted in the total time just as they would in a paper assessment.

In addition to a reading-time measurement, the WCPM calculation also requires a count of words correctly read. In a paper assessment, words can only be ‘correct’, or ‘incorrect’, whereas in Reading Assistant™ there are three levels of color coding for correctness level: red, blue, and green. Red indicates a word where the user was given a full audio intervention and was prompted with the word; blue coding typically indicates a more subtle error, stumble, or hesitation on a word; green indicates a correctly read word. For the purposes of the words-correct count, red-coded words will not be included in the words-correct total, but all other words will be included.

The Words Correct Per Minute algorithm described above was implemented for Reading Assistant Expanded Edition. To test the WCPM calculation and whether it accomplished the goal of reducing the ‘time penalty’ caused by interventions, we compared manually determined WCPM to Reading Assistant Version 4 WCPM measurements, and to the new Reading Assistant Expanded Edition WCPM measurements. This comparison used 28 passages from 20 test speakers in the Southern model data collection. The passages and speakers were chosen to be representative of different age ranges, genders, and intervention rates. The passages chosen were short so that the manual editing and calculations could be done reasonably quickly. A single audio file for each passage was generated with ‘beep’ sounds inserted where interventions were given to the user. The manual total time calculation was made by ‘editing’ out repeated portions of audio where the repetition was prompted by an intervention. The manual process also involved counting the reading errors that the user made so that these could be subtracted from the total words in the passage to get a ‘words correct’ count. From these two measurements (total time and total number of words correct) a WCPM measurement was made.

<i><b>WCPM Calculation</b></i>	<i><b>Average WCPM</b></i>	<i><b>Average % Difference in WCPM Compared to Manual</b></i>
Reading Assistant Version 4	104	-13%
Reading Assistant Expanded Edition	112	-6%
Manual	118	-

*Table 5. Comparison of Reading Assistant and Manual WCPM Calculations*

Table 5 shows the results for the manual calculations on recorded data as compared to Reading Assistant™ Expanded Edition and Reading Assistant Version 4. Both the old and new Reading Assistant algorithms result in measurements that are lower, on average, than the manual editing calculations. However the Reading Assistant Expanded Edition result is much closer to the manual calculation, indicating that the new implementation is successful in significantly reducing the ‘time penalty’ caused by interventions in cases where the user is reading fairly fluently and continues reading. A gap in WCPM calculations still exists in comparison to the manual calculations. This is mainly due to recognizer errors that prevent the software from detecting that a user has read on and therefore the time correction is not made. The logic for doing the time correction can be improved based on an analysis of remaining discrepancies in future versions, thereby reducing the gap further.

### **Default Glue Word List Changes**

In Reading Assistant, words in a story are given a category assignment which influences how they are processed and treated in Record My Reading Mode. This assignment can affect logic and settings in the speech recognition software and in the associated processing and feedback modules. ‘Glue’ words are a category typically composed of short, very common words such as articles and prepositions, which it is assumed that the

reader already knows. These words are typically not important to meaning, and they are often elided or de-emphasized in speech which results in their being misrecognized. For these reasons, we do not require correct recognition of them to allow the user to proceed in the text.

Previous versions of Reading Assistant™ used a relatively long list of 65 glue words as the ‘default’ glue list, which included many pronouns and other common words. Specific words in text could be moved from the ‘glue’ category into another category, but this had to be done manually on a story by story basis.

The new content in Reading Assistant Expanded Edition focuses on below-grade-level readers and has more content at lower reading levels than previous content anthologies. Due to this shift in focus it was important to develop new shorter default glue word lists for lower content levels, since many words on the existing longer list will not be known or automatic for readers at these levels. At the same time, these changes need to be made carefully so that the false negative rate of Reading Assistant (how often we stop a user when they haven’t made an error) does not increase too much and impact usability.

New lists were developed for first and second grade reading levels, taking into account both performance and pedagogical considerations. Performance (false negative rate) was measured using automated Reading Assistant testing on recorded data. The data corpus used was the K-2 Data Collection. This data collection used an early version of the RAEE content and focused on below-grade-level readers. From a performance standpoint, the criterion used was that a reduced list could not increase the false negative rate by more than 10% relative (e.g. a 1.0% rate could not increase to more than 1.1%), relative to the original list, when tested on the subset of story readings at the target grade level or levels.

From a pedagogical standpoint, many pronouns were removed from the reduced lists since these are often the subject or object and can be important for meaning. Three longer, more difficult (in terms of letter to sound rules) words were also removed from the original 65-word list based on pedagogical recommendations.

The default glue word lists by reading level are shown in the tables below. These are used to automatically assign words in a text to the glue category. However this automatic assignment will be overridden in some cases, based on considerations including a word's importance in the particular context, as well how words are sequenced (for example how many glue words occur in a row in a sentence). Therefore a word which appears on a list below will not necessarily be categorized as a glue word every time it appears in texts at those reading levels.

the	in
a	an
and	at
to	on
of	but

*Table 4. Default Glue Word List for Reading Levels 1-1 and 1-2 (These will apply to all future 1-x reading level; currently there are only 1-1 and 1-2 content levels.)*



a	in	so
an	it's	the
and	its	to
as	no	too
at	not	up
be	of	with
but	on	yes
by	or	

*Table 5. Default Glue Word List for Reading Levels 2-1 and 2-2 (In the future, these will apply to all 2-x reading levels; currently there are only 2-1 and 2-2 content levels.)*

a	for	into	our	was
am	from	is	out	we
an	get	it	she	were
and	go	its	so	what
are	had	it's	that	when
as	has	may	the	who
at	have	me	them	with
be	he	my	then	yes
but	her	no	this	you
by	him	not	to	your
can	his	of	too	
did	I	on	up	
do	in	or	us	

*Table 6. Default Glue Word List for Reading Levels 3-1 and Higher*

## **Conclusion**

Reading Assistant™ Expanded Edition incorporates a number of significant improvements which impact its Reading Verification capability ('Record My Reading' mode). These improvements will make Reading Assistant Expanded Edition more effective in providing guided oral reading practice, building and measuring fluency skills, and supporting comprehension and vocabulary development.

## **Notes**

To cite this report: Scientific Learning Corporation. (2009). Reading Verification Improvements in Scientific Learning Reading Assistant™ Expanded Edition, Scientific Learning: Research Reports, 13(13): 1-18.