

MindPlay Research Document: RAPS 360

Table of Contents

| | |
|---------------------------------|----|
| Introduction | 1 |
| Program Description | 1 |
| Computer Assisted Testing | 3 |
| Research Based Testing Method | 4 |
| Phoneme Segmentation | 4 |
| Phonics | 5 |
| Comprehension | 6 |
| Word Meaning and Recognition | 7 |
| Eye Tracking | 7 |
| Fluency | 8 |
| Research and Evaluation Studies | 10 |
| Study 1 | 10 |
| Study 2 | 11 |
| Study 3 | 12 |
| Study 4 | 15 |
| Summary and Conclusion | 16 |

Scientifically Sound – Reality Driven

Introduction

Founded in 1981, Methods & Solutions Inc., dba MindPlay. Inc., is both a publishing and development company. Working collaboratively with dedicated teachers, researchers, and school administrators, MindPlay translates effective classroom strategies into fully functional, technology-based reading solutions. This approach contributes to extraordinary results in meeting the everyday challenges in the classroom.

MindPlay educational programs are initiated from best practices of successful teachers, grounded in scientific research and theory, and proven in the classroom. These programs are currently in use in more than 3,000 schools across the United States, Canada, and, most recently, the United Kingdom, and both Northern and Southern Ireland.

This document contains information about the research base underlying the structure of MindPlay's Reading Analysis & Prescription System (RAPS™ 360), descriptions of the program components, and highlights of the research studies conducted on the program components.

MindPlay products have won numerous awards and recognitions including the AEP Distinguished Achievement Award, H.O.T. Award, and the Bessie Award. Most recently, RAPS 360 received the Distinguished Achievement Award given by the Association of Education Publishers' (AEP) in the Technology & New Media, Assessment Category for Grades 9-12.

PROGRAM DESCRIPTION

The MindPlay technology-based reading solution includes: *My Reading Coach™*, *FLRT™-a fluent reading trainer*, and the diagnostic reading assessment *Reading Analysis & Prescription System (RAPS™360)*. RAPS 360 was developed to efficiently determine specific reading deficiencies, strategically assign interventions, and measure progress.

RAPS 360 is a student-centered diagnostic assessment. It offers a range of appropriate assessments grounded in the science of learning. By measuring silent fluency and comprehension, RAPS 360 returns the teacher's focus back to the student and quickly provides diagnostic information to promote each student's optimal learning path through easy to use and motivational software.

The RAPS 360 screening and diagnostic test sets and its extensive reporting capabilities qualify it as an assessment for learning. The goal of RAPS 360 testing sets is to identify students' strengths and weaknesses and provide the teacher with instructional guidance.

RAPS 360 requires only a minimal amount of student and teacher time. It is affordable when compared to teacher-administered assessments, and engages students in the tasks so that their

responses are more indicative of their reading ability (McBride, J. R., Ysseldyke, J., Milone, M., & Stickney, E., 2010). When combined with the program's ability to reference instructional materials, RAPS 360 is student-centered. It provides the teacher with a tool to assign the correct intervention, which will give any students the opportunity to learn and meet their full potential.

Teachers need access to meaningful assessments as well as instructional materials that are appropriate for a broad range of students. Student behaviors are measured adaptively in RAPS 360. A traditional assessment, which has a predetermined ceiling or floor, cannot gather as much information about a learner as an adaptive test can. Assessment should be used to influence learning. "Teaching should be contingent on what students have learnt, so that while we're teaching, we collect evidence about where the students are to make adjustments to our teaching to better meet our students' learning needs" (Valencia et al, 2010).

The reports generated by RAPS 360 are both informative and adaptable, allowing educators to evaluate the learning status of individual students, classes, schools, or even a district. Diagnostics to promote learning is the principle that guided the development of the reporting system. Through automatic reporting, RAPS 360 will identify underperforming students, group students who need comparable instruction, and create individual learning plans.

The assessments that are part of RAPS 360 are used for progress monitoring as well as screening. Most importantly, RAPS 360 can help the teacher assign the appropriate intervention, evaluate the effectiveness of instructional practices, and monitor the effectiveness of any assigned intervention. RAPS 360 is objective and efficient.

In April 2000, The National Reading Panel (NRP) published a report titled "*Teaching Children to Read: An Evidence-Based Assessment of Scientific Research on Literature on Reading and Its implications for Reading Instruction.*" This report concluded that there were five key components in a successful reading program; RAPS 360 diagnoses **all five of the critical reading components**: Phonemic Awareness, Phonics, Vocabulary, Comprehension, and Reading Fluency with connected text. RAPS 360 is **100%** computerized. Its tests are adaptive based on student responses. This feature allows even the most struggling readers to be diagnosed.

RAPS 360 tests are NOT administered by the teacher. Therefore, the results are unbiased, and teacher time is minimized. Data is automatically collected as students are tested so that progress can be instantly reported. The program tracks students' growth and the effectiveness of any assigned intervention. Administrators and teachers will know which interventions are working with which types of students.

RAPS 360 measures meaningful reading behaviors. RAPS 360 currently includes eight diagnostic tests: the Comprehension Screening Assessment, Phoneme Segmentation, Word Meaning and Recognition Assessment, Phonics-Decoding/Encoding Assessment, Eye Tracking Assessment, Natural Fluency Assessment, Expected Fluency Screening, and the Pause-Assisted Fluency Assessment (Reading Connected Text Assessment). These assessments are grouped into three test sets that provide either detailed diagnostics or a quick progress check. Depending on the test set and test performance, students may take as few as one or as many as five assessment and diagnostic tests.

COMPUTER ASSISTED TESTING

Computer assisted testing is perhaps the most efficient way to conduct appropriate assessments in a cost-effective way without reducing instructional time. The use of computerized testing should be an attractive alternative to recognized early reading measures (McBride, J. R., Ysseldyke, J., Milone, M., & Stickney, E., 2010).

The computer algorithm in adaptive testing allows the student's level of achievement to be analyzed after each item and then uses the information to assign subsequent items. In addition, computer adaptive testing (CAT) can yield reliability and validity comparable to traditional paper-and-pencil test with far fewer items (Wainer, 2000).

Most paper-and-pencil tests require all students to answer the same questions within a given test. Since everyone takes every item, all students are subjected to items that can be either very easy or very difficult for them. As a result, the test may give little information about the student except that they "don't know" grade level material. Computer adaptive tests (CAT) can provide accurate scores over a wide range of abilities while traditional tests are usually most accurate for average examinees (Grist, 1989).

The CAT item selection helps reduce standard errors with greater accuracy and with fewer items needing to be completed (Rudner, 1998). The RAPS 360 CAT algorithm follows a standard CAT algorithm and uses the following steps:

1. All the items that have not yet been administered are evaluated to determine which will be the best one to administer next given the currently estimated ability level.
2. The "best" next item is administered.
3. A new ability estimate is computed based on the response.
4. Steps 1 through 3 are repeated until a stopping criterion is met.

Computerized testing provides these benefits that a paper and pencil tests cannot:

- Computerized tests can be given on an "as needed" basis.
- Scoring data is available immediately.
- Testing protocol and security is consistent for all classes and students. Teachers may not preview a test that will be given.
- Adaptive testing differentiates for all students and allows extra time for students who need it.
- Less time is needed to administer a CAT, giving more teaching time back to the teacher.
- Adaptive tests and shorter testing times help reduce testing stress of many struggling learners.

Scientifically-Based Research

TESTING METHOD- RAPS 360 Testing Components

A Full Diagnostic Assessment is automatically scheduled by benchmark 3 times per year. Although there are eight possible adaptive tests, the student's path through testing depends on his/her skills.

A Progress Monitoring Assessment is automatically scheduled every 14 days. The progress monitoring schedule is changeable through the management system. The purpose of progress monitoring is to give teachers a frequent and quick way to monitor student progress in BOTH comprehension and fluency.

RAPS 360 tests the following critical reading components:

PHONEME SEGMENTATION FLUENCY (K-1)

Phonemic awareness refers to the ability to focus on and manipulate phonemes in spoken words. This adaptive test was developed to screen for students who may need phonemic awareness instruction. The test uses an activity called "Counting Sounds" which evaluates their phoneme segmentation abilities.

Conclusions from the Report of the National Reading Panel (2000) suggest that phonemic awareness instruction helps all children including normally developing readers in kindergarten or first grade, readers with disabilities, preschoolers, and other students with reading disabilities. Furthermore, "PA measured at the beginning of kindergarten is one of the two best predictors of how well children will learn to read" (NRP, 2000, p. 2-11).

The Phoneme Segmentation Test quickly identifies students who would benefit from direct phonemic awareness instruction. This RAPS 360 test gives the teacher a tool to begin the year knowing who is ready to enter into phonics instruction, or who needs further work with phonemic awareness. The teacher no longer needs to wait to see what particular skills (sounds) a student knows.

The Phonemic Awareness Test is a special and unique test for students in kindergarten and first grade. Students in these grades work on pre-reading activities and are not expected to be reading passages at this point in their development. However, students should have the ability to correctly count the number of sounds they hear in a word.

This test uses a virtual coach to say a word to the student. The student then clicks a "token" and slides it to the top. He can only slide one at a time. The total number of tokens at the top should be the total number of sounds in the word given by the coach.

The words used in the test are grouped into six levels to test from 2 – 7 sounds in progressively longer words. Each higher level has words with concepts from the previous levels.

PHONICS (for levels kindergarten – 1st grade)

This Adaptive Phonics Test was developed to screen students who may need phonics or phonetic spelling rule instruction. RAPS 360 tests a student's ability to correctly hear sounds and spell words. Basic sound/letter correspondence and short vowel single syllable words are assessed. Advanced students may progress further.

PHONICS (for levels 2nd grade - adult)

The Phonics Test is the final test given in the Full Diagnostic Test set for grade 2 and above. The purpose of this diagnostic test is to estimate the student's decoding ability and determine if there are gaps that are affecting his/her comprehension.

This adaptive test was developed to screen students who may not be reading fluently because they lack decoding automaticity. These are students who may need phonics or phonetic spelling rule instruction. Up to nine skill levels are accessed. Students enter the program at a level appropriate to their grade level. The program quickly adapts the difficulty level based on student performance.

The ultimate goal of reading is comprehension (NRP, 2000). However, before a student can get to comprehension they must be able to decoded words with automaticity (Kuhn & Stahl, 2008). The report of the National Reading Panel concluded that phonics instruction is one of the five critical requirements for effective reading instruction. According to Carnine, D., Silbert, J., Kame'ennui, E.J., Tarver, S., & Jungjohann, K. (2006), there are six basic approaches to teaching phonics. They suggest using a phonics approach which focuses on teaching students phonics though spelling where students learn to pronounce sounds in isolation, to convert letters or letter combinations into sounds, and to then blend the sounds to form words.

The RAPS 360 phonics test analyzes a student's knowledge of English letter/sound correspondence, sound-spelling, roots and spelling rules. The words used in the test are grouped into nine levels to test a progression of more complicated phonics sounds and concepts. Each higher level has words with concepts from the previous levels. The following is a list of the phonics skill levels.

Phonics Skill Levels:

1. Sound-letter Correspondence
2. Short Vowels and Consonants
3. Short Vowels, Consonants, Digraphs and Blends
4. Long Vowels w/silent e
5. Long Vowel Combinations: ea, oa, ai, ay, oo, aw
6. Long Vowel Combinations and diphthongs: oi, oy, ow, ou, au, aw
7. Rules, R-Controlled, FLOSS
8. 2-syllable, open/closed syllables, suffix endings
9. Multi-syllable, prefix, suffix, syllabication rules, Greek roots

The Phonics Test uses a virtual teacher to say a target nonsense word. Nonsense words are used to accurately test a student's decoding ability. It prevents the student from selecting the correct answer only from familiarity with an actual word. The auditory presentation by a virtual teacher helps the student visually distinguish similar sounding letters. The student chooses the one correct word from a list of five. The other four choices are special distracter words. Each distracter word tests at least one common decoding error (such as audio or visual discrimination). This makes any student errors more meaningful beyond just not selecting the correct answer.

The Woodcock Reading Mastery Test – Revised (WRMT-R) utilizes nonsense words with the implication that the use of phonetic nonsense words and syllables allows the examiner to determine how the child recognizes words without using other strategies such as looking at pictures, guessing based upon context, or reading words by sight (Farrall, 2008)

COMPREHENSION (for levels 2nd grade – adult)

This adaptive screen was developed to test students' reading comprehension level quickly and effectively. This one to two sentence modified cloze test is administered by grade level. Students entered into the program with special learning requirements are not subjected to a long passage beyond their ability. Yet, their highest possible independent reading level is determined. Students who pass this test at or above grade level, progress to a more comprehensive reading test.

Comprehension Screen– Cloze Test

The initial Comprehension Screen (CSA) is a modified cloze test. It is adaptive based on student responses. Students with good comprehension proceed to a series of three different silent reading fluency tests due to the importance of determining a student's potential obstacles preventing fluency. Students who cannot pass at the initial fluency test (Natural Fluency) at the desired rate for his/her grade level are further analyzed to determine how teachers can assist them. The additional tests can uncover difficulties in attention, grammar, eye-tracking, and decoding.

The RAPS 360 (CSA – comprehension screening assessment) test reflects two major reading skills: vocabulary knowledge and the ability to derive word meaning from context. A **cloze test** is a test in which the student reads text containing a blank and chooses the best word to fit in the blank. Words may be deleted from the text in question either mechanically (every nth word) or selectively (modified). The RAPS 360 modified cloze is designed to test the meaning of words that are intrinsic to the meaning of the sentence.

For group or computer-based assessment, there is no procedure that produces more dependable information than the cloze test (DeVito, 1967). A further advantage is that this assessment task is relatively culture-fair and does not depend on family status and educational background to the same extent, as do other assessment procedures (He, 2002).

The target words on which the assessment is based were drawn from a composite list that reflects a number of classic research-based word frequency lists. They include:

- The Educator's Word Frequency Guide
- The American Heritage Word Frequency Book

- The Living Word Vocabulary
- The Harris-Jacobson Basic Reading Vocabularies

WORD MEANING AND RECOGNITION DIAGNOSTIC TEST

This is a unique assessment that tests vocabulary without reading comprehension variable. The oral prompting sets it apart from other assessments and makes it much more versatile. It has a very high ceiling and a low floor. This test helps determine if word comprehension is a contributing factor to below average comprehension skills.

The words selected for this test are taken directly from student errors on their previous reading comprehension screen (CSA-COMP). Student results are graphed against goal. At this time, fluency tests are not administered because they could be frustrating to these students.

Students who do NOT pass the initial comprehension screen (CSA-COMP) proceed to the Word Meaning and Recognition Test. This test ensures that the vocabulary delivered in the CSA-COMP passages did not prevent student success.

In the Word Meaning and Recognition Test, RAPS 360 uses words from the student's most recent Comprehension Screening Test to determine if the student did not pass the screening test because they were struggling with vocabulary. The vocabulary is spoken to ensure the student did not stumble in the comprehension screening because he did not understand the vocabulary.

The meaning prompt is auditory and the answers are printed.

Which of these is a color? (spoken)

and lie red (NOT spoken)

The auditory presentation of the prompt makes the WMRT a pure measure of vocabulary knowledge. If the student does well on the WMRT, then it is likely that the comprehension problem is related to fluency, vision, decoding, or processing difficulty. Other assessments will help to identify which of these difficulties is present.

The Word Meaning and Recognition Test is a context free assessment of a student's ability to match a printed word with its auditory representation. If the student knows the meaning of the word, the student will choose the correct answer.

EYE TRACKING

This test helps determine if poor eye tracking or eye-teaming is a contributing factor to below average comprehension skills. It can determine if inefficient eye movement is affecting reading rate. Student results are graphed against goal. Students who do not meet the grade level goal are assigned the Eye-Tracking Test. Students who meet the grade level goal for comprehension and fluency will not be assigned the Eye-Tracking Test.

This Eye-Tracking Test determines if the student’s eyes properly team together and flow across the page, left to right and top to bottom. Shapes or numbers move across the screen at a certain rate, with only one shape visible at a time. The shape seen most of the time will be the alternate shape. During the course of the test, the target shape will appear instead of the alternate shape.

The student is asked to count and record the number of times the target shape appears. The Eye-Tracking Test identifies any potential eye-function issues. Students track from left to right across the page counting target shapes at a rate appropriate for their grade level. This test can identify potential functional eye problems like eye-teaming that may be preventing a student from achieving fluency.

Symptoms such as letter reversals, headaches, eye-strain, fatigue, printing backwards, skipping words, seeing double, letters or words moving on a page, are some of the challenges students face when their eyes do not focus or team together properly (Crane A & Crane V, Davis, M., Wick B, 2009).

The Orfield & Basa study (2001) tested schoolchildren’s vision skills and concluded, “Without counting the visual tracking test, 41% failed; adding the tracking test, 53% failed. The Developmental Eye Movement Test, which 24.5% of the children failed, is an excellent predictor of a significant percentage of reading failure.”

Other research suggests that eye movement difficulties are associated with reading problems, yet few assessments measure this aspect of perception (Hutzler, F. & Wimmer, H., 2004)

FLUENCY - Reading Connected Text Comprehension

Once the student passes the Comprehension Screening Test at or above grade level, he moves onto fluency testing. The three fluency tests (Natural, Expected, and Pause-assisted) are given to diagnose the student’s difficulty with fluency. They read a grade-appropriate passage at their own rate. Their performance on both comprehension and fluency is compared against goal for their grade level.

Students who do not meet goal in Natural Fluency are given a different test (Expected Fluency) and passage where the words fade out at their current grade-level goal rate. This masking technique forces students to read at the appropriate word rate.

There is a close relationship between fluency and comprehension (NRP, 2000). However, commonly used oral fluency rates usually only calculate words per minute and do not correlate the rate with a comprehension score. RAPS 360 calculates a student’s “effective reading rate” – the highest rate a student can read while maintaining a minimum comprehension score of 80%. RAPS 360 fluency wpm rates fall into the median ranges calculated by Harris and Sipay (1990) from 6-8 different standardized reading tests (p.634). A single word per minute scores cannot provide the depth of information needed to make good instructional decisions or programmatic recommendations. “The wcpm metric itself obscures the relationship between accuracy and

automaticity, and it fails to capture other critical aspects of skilled reading (Valencia et al, 2010).”

In contrast, RAPS 360 measures actual reading behaviors, and because of the variety of assessments, informs instruction in specific ways. Moreover, several of the RAPS 360 assessments are unique and offer perspectives that are not otherwise available.

Researchers and educators alike agree that comprehension is the ultimate goal of reading. If assessments and their subtests that are currently used in schools are not closely related to comprehension, misallocation and inappropriate assigning of resources may occur (Valencia, S. et al 2010).

Natural Fluency (Reading Connected Text)

This fluency screening assessment presents a multiple-line story. Students are encouraged to read the text as quickly and accurately as possible. After the story has been read, questions about the story will be given to verify that the student understood what was read. This test is timed and if the student is reading on grade level at the speed appropriate to that grade level, the student meets benchmark and testing is complete.

Expected Fluency (Reading Connected Text)

This fluency screening assessment presents a multiple line story at a specific reading rate. Multiple lines will be visible one at a time. Beginning at the top of the page, the words will fade out at the predetermined rate. This masking determines if the student has the ability to read at the appropriate word rate and scan down the text in addition to scanning left to right with a minimum amount of word fixations and regressions. After the story has been read, questions about the story will be given to verify that the student understood what was read. Students who pass Expected Fluency but fail Natural Fluency have the skills to read fluently but may lack attention or motivation.

Pause-Assisted Fluency (Reading Connected Text)

This test determines if a student is able to read more fluently when words are ‘chunked’ in a story. For the non-fluent reader who reads word by word, the technique of chunking a group of meaningful words together by natural pause points helps students internalize and gain meaning from what is read. Chunking encourages fluent reading by grouping word phrases and enabling the comprehension of complete thoughts.

The paused-assisted RAPS 360 provides additional diagnostic information for a teacher to get a better idea of why a student may experience difficulty reading fluently. Teachers who assign an ‘assisted reading’ intervention will provide learners with a correct model of fluent reading (Kuhn & Stahl, 2008). Students who pass Pause-Assisted Fluency but fail Natural or Expected Fluency would benefit from grammar instruction – specifically clauses, phrases, prepositions and conjunctions.

RESEARCH AND EVALUATION STUDIES

In 2007, MindPlay began a rigorous research program to systematically obtain reliable and valid knowledge about RAPS 360 and the impact it has in the classroom. Highlights from several studies are presented here for review.

Study 1

As part of a larger evaluation study of *My Reading Coach* during the 2007-08 school year, an analysis of the reliability and validity of RAPS was conducted. The criterion instrument for the analysis was the Metropolitan Achievement Test[®], Eighth Edition (MAT-8). (Pearson Assessment, 2000.)

At the beginning of the evaluation study, both RAPS and the MAT-8 were administered to participating students in grades two, three, and nine in public schools in Arizona and Texas. The MAT-8 was administered to students as a group by the teacher. RAPS was administered individually through the computer. A second administration of RAPS took place several weeks later for the purpose of test-retest reliability. RAPS scores were percent correct, and MAT-8 scores were scaled scores. The Total RAPS score is the average of the percent correct for the subtests.

The total number of students who participated in the beginning of the study was 501. The grade distributions are shown below.

| Grade | Number of Students |
|-------|--------------------|
| 2 | 284 |
| 3 | 98 |
| 9 | 119 |

The pretest mean for Total RAPS was 70.49 with a standard deviation of 19.27. The mean for the second administration was 75.14 with a standard deviation of 18.23. The increase in the mean from one test to another is expected because students received direct instruction in relevant reading skills. **The correlation between the first and second administrations of RAPS was 0.81. This meets the reliability coefficient level recommended by Salvia and Ysseldyke (2008).**

The concurrent validity correlations of RAPS percent scores with the MAT-8 scaled scores are shown below. Although these two instruments do not measure identical skills, they assess behaviors that are closely associated with reading success. **The majority of the coefficients are above 0.50, which is a widely accepted criterion for concurrent validity (Peers, 1996).**

The correlation coefficients in the table below range from 0.15 to 0.77. Of the RAPS subtests, long vowel recognition has the most consistently high correlations with the MAT-8 subtests and total score. The Total RAPS score correlated adequately with all of the MAT-8 subtests and total score.

The Structure subtest of RAPS had the lowest correlation with the MAT-8. This low correlation suggests that the subtest measures a skill that is relatively independent of those tested on the MAT-8. The Structure subtest is an encoding task that measures an aspect of literacy that does not lend itself to the multiple choice format of the MAT-8.

| MAT Subtest* | Sounds | Language | Vocabulary | Comprehension | Total Reading | Spelling |
|---------------------|---------------|-----------------|-------------------|----------------------|----------------------|-----------------|
| Total RAPS | 0.54 | 0.55 | 0.51 | 0.52 | 0.56 | 0.50 |
| Short Vowels | 0.52 | 0.46 | 0.39 | 0.34 | 0.40 | 0.35 |
| Long Vowels | 0.69 | 0.72 | 0.71 | 0.71 | 0.77 | 0.76 |
| Consonants | 0.57 | 0.51 | 0.44 | 0.41 | 0.45 | 0.42 |
| Blends | 0.58 | 0.55 | 0.54 | 0.53 | 0.56 | 0.51 |
| Structure | 0.29 | 0.25 | 0.15 | 0.19 | 0.18 | 0.15 |
| Rules | 0.43 | 0.58 | 0.59 | 0.62 | 0.66 | 0.61 |

*The MAT subtests are pretest scores on recognizing sound spellings, vocabulary knowledge, reading comprehension, total reading, and spelling.

References

Peers, Ian (1996) *Statistical Analysis for Education and Psychology Researchers*. New York: Routledge

Salvia, John & Ysseldyke, James. (2000) *Assessment* (8th edition) Boston: Houghton Mifflin

Study 2

One way of determining the validity of a measure is to see how well it corresponds to the expected behavior of the population for which it was intended. This approach uses modeling, a research innovation described by Joseph Rogers (2010), who believes that rather than traditional null hypothesis significance testing, "...researchers and textbook authors should be discussing how to create and compare behavioral models that are represented mathematically and evaluated statistically (p. 1)."

The model of behavior in this case is the general increase in reading ability over the grades. An assessment should be sensitive to this change to the degree to which it measures various behaviors. An assessment of letter knowledge, which is a constrained skill, should reach a ceiling relatively quickly, after which no improvement is possible. In contrast, comprehension is an unconstrained skill, and it may continue to improve as long as the difficulty of the text on which it is based increases. Eventually, however, except in extreme cases, a ceiling will be reached, after which a plateau or even decline may be experienced.

RAPS 360 was used with 496 students in grades 1 through 8 of a charter school. The two measures used for this study were a modified-cloze comprehension screening assessment and a silent fluency assessment. The scores returned for the comprehension screening assessment is a grade level, and for the fluency assessment, words per minute.

The median grade level of the students was 5, and the median score for the comprehension assessment was also 5. The correlation between grade and comprehension screening was 0.62. Both of these statistics support the traditional model of educational progress in which students' reading comprehension ability is more or less represented by their grade level.

A subset of the students (N = 314) was also administered the silent fluency assessment to estimate their words correct per minute reading speed. A literal comprehension assessment followed the silent reading to ensure that the students had some understanding of what they read and were not simply skimming the text. The correlation between the comprehension screening and fluency scores was 0.37, a moderate correspondence.

This moderate correlation is consistent with the belief of many reading researchers that there is an "imperfect" correlation between fluency and comprehension. Some fluent readers will not comprehend very well, while some who are disfluent will manage to understand text anyway (Shanahan, 2009).

Given the degree to which RAPS 360 represents the distribution of comprehension scores and the correlation between comprehension and fluency scores, it is safe to assume that it is a valid measure of these two reading behaviors. A limitation of this study is that no outside assessment was used as a criterion. Further research was needed to overcome this limitation.

References

Rogers, Joseph Lee (2010) The Epistemology of Mathematical and Statistical Modeling: A Quiet Methodological Revolution. *American Psychologist*, 65, 1-12.

Shanahan, T. (2009) Demolishing a Straw Man: Should We Teach Fluency?
<http://www.shanahanonliteracy.com/2009/03/demolishing-straw-man-should-we-teach.html>

Study 3

As part of our ongoing research efforts, RAPS 360 was compared against several other measures in order to evaluate its validity. The term *validity* describes how well an assessment measures the intended construct. Previous reports have supported both the validity and reliability of RAPS 360, but the current studies expands the scope of the criterion assessments against which it was compared.

The RAPS 360 screening process produces a measure called a Composite score (Comp). This score is based on students' performance on a cloze-type comprehension screening assessment, a measure of silent reading speed, and their responses to comprehension questions asked about the fluency passages.

SRI® and RAPS 360

The Scholastic Reading Inventory (SRI®) is a computer-adaptive test of reading comprehension. It uses the modified-cloze format and generates a Lexile® score. It was chosen as a criterion assessment because the SRI is similar to the screening assessment component of RAPS 360 in its item format and administration procedure. The participants in this study were 201 students in grades 4 and 5 in an urban school in Arizona. School profile: large urban school district, Tucson, AZ 655 students; Grades K–5; 92% Hispanic; 33% ELL; 85% Title 1.

The overall correlation between the RAPS 360 score and the SRI was 0.68, suggesting that the two tests are measuring the same underlying construct. The correlation was slightly higher for the 94 students in grade 4 compared to the 107 students in grade 5.

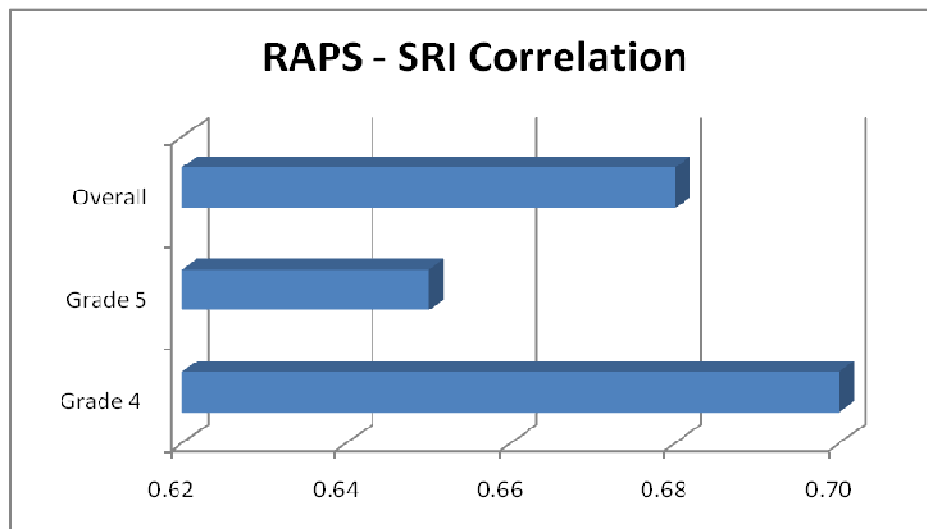


Figure 1. Correlation between RAPS 360 and SRI Overall and by Grade

RAPS 360 was a more conservative test, generating a score that is approximately half a grade level below the SRI score. This discrepancy is not surprising, as the RAPS 360 score comprises two measures of comprehension and a fluency measure, while the SRI tests comprehension using just one item format. In addition, RAPS 360 was designed as a screening and progress monitoring assessment intended to identify students at risk of reading failure. The SRI, in contrast, was initially developed to match students with books at their reading level, so it may be less able to identify students who are underperforming.

DIBELS® and RAPS 360

The Dynamic Indicators of Basic Early Literacy Skills (DIBELS®) is a teacher-administered test of several reading behaviors. The most common of these is oral reading fluency, which chiefly measures how well students read text out loud. This instrument was chosen because it is widely used and its administration procedure is markedly different from RAPS 360. The teacher and student being assessed must move to a dedicated assessment area while the rest of the class does

seatwork or is directed by a teaching assistant. Despite its popularity, DIBELS is a relatively expensive assessment to administer because of the cost associated with teacher time needed for the testing process and the need for an assistant or substitute.

This study was also conducted in the same urban school in Arizona. The participants in this study were 104 students in grade 3. School profile: large urban school district, Tucson, AZ 655 students; Grades K–5; 92% Hispanic; 33% ELL; 85% Title 1.

The traditional use of DIBELS is to categorize students into three groups that reflect the likelihood of reading below the expected level for their age or grade: at risk, some risk, and low risk. In this study, we examined how closely RAPS 360 matched the DIBELS categorization.

The overall percentage match between the two measures was 88%. A summary by grade of students' DIBELS category and placement by RAPS 360 is presented in Table 1.

Table 1. Correctly Matched by RAPS 360 based on DIBELS Category

| | Students by DIBELS Category | Students Correctly placed by RAPS 360 | Percent Correctly placed by RAPS 360 |
|-----------|-----------------------------|---------------------------------------|--------------------------------------|
| At risk | 41 | 36 | 89% |
| Some risk | 33 | 29 | 88% |
| Low risk | 30 | 9 | 30% |

In the at risk and some risk groups, there was a high degree of overlap, with RAPS 360 matching the DIBELS categorization for 89% and 88% of the students respectively. The correspondence was less in the low risk group, where the RAPS 360 to DIBELS match was only 30%. This finding is not unexpected because DIBELS relies primarily on oral reading speed, while RAPS 360 uses two measures of comprehension as well as silent reading speed. These results are depicted graphically in Figure 2.

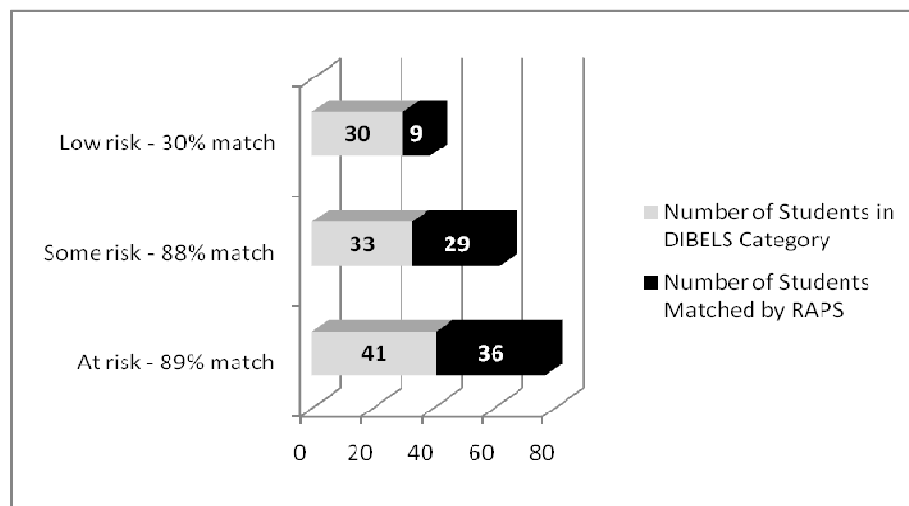


Figure 2. Number of students by DIBELS category and percentage match to RAPS 360

Study 4

This study was conducted with 335 students in grades 3 through 8 in an inner-city school in Pennsylvania. School Profile : 600 students; Grades Pre-K–8; 100% African American; 72% Disadvantaged; Title 1.

PSSA and RAPS 360

The criterion instrument was the Pennsylvania System of School Assessment (PSSA). For this predictive validity study, we examined how well students' RAPS 360 scores predicted their performance on the PSSA. A summary of the results is presented in Table 2.

Table 2. Correctly Matched by RAPS based on PSSA Category

| | Number of Students in PSSA Category | Number of Students Matched by RAPS 360 | Percent Identified by RAPS 360 |
|-------------|-------------------------------------|--|--------------------------------|
| Below Basic | 77 | 70 | 91% |
| Basic | 71 | 63 | 89% |
| Proficient | 128 | 57 | 45% |
| Advanced | 66 | 38 | 58% |

From the point of view of teachers and administrators, one of the most critical aspects of an assessment is its ability to predict students who will not achieve proficiency on the state test. These students would be most likely to benefit from an intervention, which in turn will reflect favorably upon the overall rating of the school. For this purpose, RAPS 360 functioned exceptionally well (see Figure 3). It identified 91% of the students who were categorized Below Basic and 89% of the students who were at the Basic level. For students at the Proficient and Advanced levels, RAPS 360 was not as dependable. It identified 45% and 58% of the students in these categories respectively. A summary of these results is displayed graphically in Figure 3.

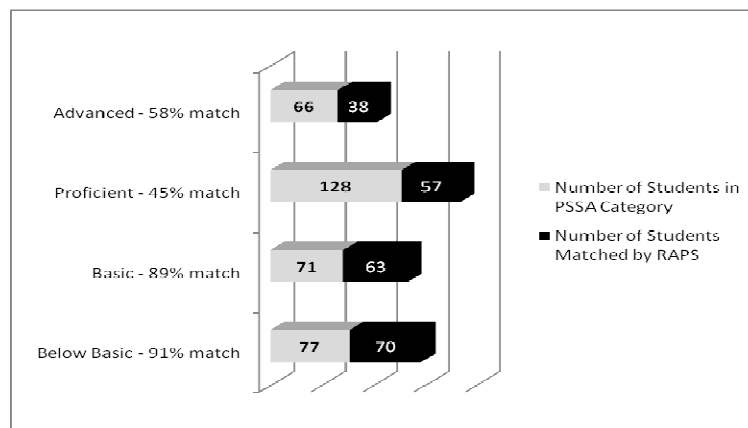


Figure 3. Number of students by PSSA category and percentage match to RAPS

This discrepancy suggests that RAPS 360 underestimates the ability of better readers because it is more challenging than the PSSA. Of the miscategorized students, the majority (84%) were classified just one year below grade level. In short, RAPS 360 is a conservative measure of Proficient and Advanced students' ability as measured by the PSSA. The consequence of this underestimation is that some students who are at the Proficient and Advanced levels would receive somewhat more intensive instruction than they might need.

Summary and Conclusion

In the four studies described here, RAPS 360 was found to correspond well to other measures of reading ability. The correlation with the Scholastic Reading Inventory(SRI) ranged from 0.65 to 0.70, which can be described as moderately high. When used to categorize students as being at risk or at some risk, RAPS 360 agreed with DIBELS for 88% of the students tested. In the study that was based on the most rigorous criterion, the Pennsylvania System of Student Assessment, RAPS 360 matched the PSSA exceptionally well, successfully predicting 91% of the students who were categorized as Below Basic and 89% of the students who were at the Basic level.

The major discrepancies between RAPS 360 and the other instruments was in identifying students who were scoring at or above grade level. In all three cases, RAPS 360 generated lower overall scores, suggesting that it is a more difficult assessment. It appears that this is partly the result of the three sources of information that constitute the Comp score, a cloze-based comprehension screening, a measure of silent reading fluency, and comprehension questions asked about the fluency passage.

As part of the ongoing research program for RAPS 360, we will continue to analyze new and existing data in an effort to more accurately identify students who are reading at or above grade level. These efforts will include additional item calibrations and an evaluation of the silent reading speed criteria we have established. Our goal is to maintain the ability of RAPS 360 to identify students who are having reading difficulties while more accurately predicting which students have superior reading abilities.

References

- Carnine, D., Silbert, J., Kame'ennui, E.J., Tarver, S., & Jungjohann, K. (2006). *Teaching Struggling and At-Risk Reader – A Direct Instruction Approach*. NJ: Pearson Education, Inc.
- Crane A & Crane V, Davis, M., Wick B. *Lions Crane Literacy Program Handbook: Response to Intervention Made Easy* NM Lions Crane Reading Foundation , 2nd Printing September 2009
- Crane A, Wick B. Preliminary assessment of the relationship between classroom symptoms and eye movements measured using the Visagraph II. *OptomVision Sci (Suppl)* 1999 Dec; 76:131
- DeVito, J. (1967) Cloze procedure. *Communication Quarterly*, 15, p. 31 - 32.

Farrall, M. (2008). Reading Tests: What They Measure, and Don't Measure. (Online Wrightslaw), August 24, 2010, <http://www.wrightslaw.com/info/test.read.farrall.htm>

Grist, S. (1989). Computerized Adaptive Tests. (Online ERIC Digest), July 29, 2010, <http://www.ericdigest.org/pre-9213/tests.htm>

Harris, A.J. & Sipay, E.R. (1990). *How to Increase Reading Ability – A Guide to Developmental & Remedial Methods*. New York and London: Longman.

He, L. (2002). Test Taker Characteristics and Performance on a Cloze Test: An Investigation Using DIF Methodology. Master's Thesis, The University of British Columbia.

Hutzler, F. & Wimmer, H. (2004). Eye Movements of Dyslexic Children When Reading in a Regular Orthography. *Brain and Language*, 89, 235–242.

Kuhn, M.R. & Stahl, S.A. (2008). Fluency: A Review of Development and Remedial Practices. (Online Ciera), July 28, 2010, <http://www.ciera.org/library/reports/inquiry-2/2-008/2-008.html>

McBride, J. R., Ysseldyke, J., Milone, M., & Stickney, E. (2010). *Technical adequacy and cost benefit of four measures of early literacy*. *Canadian Journal of School Psychology*, 25, 189– 20.

National Institute of Child Health and Human Development. (2000). Report of the National Reading Panel. Teaching children to read: An evidenced-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups (NIH Publication No. 00-4754). Washington, DC: U.S. Government Printing Office.

Orfield, A., Basa, F., Vision Problems in an Urban School Clinic: Their Epidemic Numbers, Impact on Learning, and Approaches to Remediation, *J Opt Vis Dev*, Volume 32/Fall 2001

Rudner, Lawrence M. (11/98). *Adaptive Testing, An On-line, Interactive Computer*
<http://echo.edres.org:8080/scripts/cat/catdemo.htm#whatwhy>

Serido, J & Wilhelm M.S. (2008). *Evaluation and Research Study of My Reading Coach & FLRT - A Fluent Reading Trainer 2-year report*, Arizona Board of Regents. University of Arizona. The Norton School of Family and Consumer Sciences. (Online), January 2010, www.mindplay.com/Results/Research/tabid/74/Default.aspx

Valencia, S. et al (2010) Oral Reading Fluency Assessment: Issues of Construct, Criterion, and Consequential Validity. *Reading Research Quarterly*, 45, 270-291

Wainer, H. (2000) *Computer-Adaptive Testing: A Primer*. Mahwah, NJ: Lawrence Erlbaum Associates

Williams, D. (2006) *Assessment for Learning: why, what and how*. (Online), July 1, 2010, http://web.me.com/dylanwiliam/Dylan_Wiliams_website/Papers_0409_files/Cambridge%20AFL%20keynote.doc