

# ISIP Concurrent and Predictive Validity Study

April 16, 2007

Prepared by Patricia Mathes, Ph. D.  
Professor and Texas Instruments Chair of Reading

Institute for Reading Research  
Southern Methodist University

## ISIP CONCURRENT AND PREDICTIVE VALIDITY STUDY

In this study, concurrent validity information was gathered on ISIP kindergarten, first, second, and third grade benchmark assessments compared to *Dynamic Indicators of Early Literacy Skills* kindergarten, first, second, and third grade benchmark assessments (DIBELS, Good, Kaminski, & Dill, 2002). Further, predictive validity information was gathered on ISIP Third-grade benchmarks compared to Texas Assessment of Knowledge and Skills (TAKS) at third grade. Our hypothesis was that ISIP Benchmarks will be as valid an indicator of reading skills as a well-accepted, comparable assessment that is administered one-on-one with a child (i.e., DIBELS). Specifically, our research questions were:

1. What is the alternate form reliability of each ISIP alternate form for each Benchmark subtest and for each DIBELS alternate form for each Benchmark subtest? Are the alternate form reliabilities of these two tests in line with one another?
2. What are the concurrent validity coefficients between ISIP and DIBELS for subtests measuring the same constructs?
3. Do ISIP and DIBELS identify the same children as high risk, some risk, and low risk?
4. Do ISIP and DIBELS measures given at 3<sup>rd</sup>-Grade result in similar predictive validity coefficients when compared to the Texas 3<sup>rd</sup>-Grade TAKS in reading?

ISIP benchmark assessments are designed to be administered 4 times throughout the year at approximately 2-month intervals. While ultimately the goal is to create a continuous progress monitoring system that can be administered repeatedly through the academic year using Computer Adaptive Testing, the version of ISIP under current study utilizes alternate parallel forms.

### Creation of Alternate Forms for Benchmark Assessments

Four alternate forms of ISIP were developed that would sample ability across the school year to demonstrate student progress. To develop parallel forms of each subtest, attention controlled for test administration, test length, item exposure, content (i.e., curriculum expectations for each grade level), and cognitive level (i.e., item discrimination power and level of difficulty).

**Test administration.** Tests and test items are delivered and automatically scored by ISIP, utilizing the same format and protocols for each administration and an unbiased scoring methodology.

**Test length.** Time allowed to complete each subtest is consistent across all administrations. Test length varies by skill domain and ranges from 30 seconds to 2 minutes.

**Item exposure.** The number of test items available is consistent for each skill domain and administration. The number of answered questions varies only based upon student ability to respond within the fixed assessable time.

**Content and Cognitive Level.** Item development began first with the large pool of items previously developed to measure the ISIP content areas under the direction of Drs. Mathes and Torgesen for ISIP 2.0. Applying the concepts of item response theory (IRT) models, these items were previously shown to have a high degree of discrimination validity and reliability. Items were selected from this pool based upon their discrimination and difficulty parameters and grouped according to grade level curriculum expectations as specified by state content standards, including the Texas' Essential Knowledge and Skills (TEKS).

The discrimination and difficulty parameters of these items were carefully studied. Similarly, the items used in tests of similar construct to ISIP 3.0 were reviewed. New items were created that met grade level expectations and that matched as closely as possible the parameters of the studied items. Distracters for multiple choice questions were created that followed the discrimination approach and characteristics of the studied items. Items were created for each grade level and subtest that included a range of difficulty that was consistent with the items studied. Each of the items was carefully reviewed by istation education professionals to eliminate any bias, confusion or ambiguity.

Test information and item results have been examined. First, the students were observed to gain insight into student reaction to the questions and the test overall. Next, data from student responses were evaluated. Unsatisfactory items (i.e., those that were either too easy or difficult) were replaced. The results from research and student data collected through test administration will continue to be used to refine the ISIP subtests on an ongoing basis.

## **Method**

### ***Participants***

*Schools.* This study was conducted in four elementary schools in a mid-sized (approximately 35,000 students), southwestern suburban/urban school district in the Dallas Metropolitan area of Texas. Three of these schools were identified as Recognized status, and one was Exemplary as determined by the Texas Education Agency. These schools were identified by district personnel as having student populations that were highly diverse and representative of the U.S. population as a whole.

Table 1. School demographic information

	School									
	School 1		School 2		School 3		School 4		All	
	n	%	n	%	n	%	n	%	n	%
Number of Students	395		575		530		500		2000	
Ethnicity										
African American		20.4		17.9		50.4		17.1		26.8
American Indian		0.5		0.9		0.7		0.5		0.7
Asian/Pacific Island		21.4		22.7		20.0		9.7		18.5
Caucasian		42.9		40.3		15.2		57.1		38.4
Hispanic/Latino		14.8		18.2		13.7		15.5		15.7

*Students.* Each school identified 20 students at each of 4 grade levels (i.e. kindergarten through third grade) for a total of 80 participants per school and a total sample of 320 student participants. Schools were asked to identify students within their schools that conformed to the school age population of the U.S. as a whole according to ethnicity, percentage of students on free and reduced lunch, and the number of children with special needs. Once identified, parental consent for participation was collected. We began the study with 320 participants and ended the study with 302. The 8% attrition rate reflects students who withdrew from school while the study was being conducted or for whom data could not be collected because of illness or excessive absences. Multiple attempts were made to collect all data on all participating students. Students with missing data were removed from the sample. Table 1 presents students demographic information.

Table 2: Student Demographic Information

	Grade														
	Kindergarten			1st			2nd			3rd			All		
	n	%	M	n	%	M	n	%	M	n	%	M	n	%	M
Age in years and months	75		5-11	82		7	82		8	63		9	302		7-5
Ethnicity															
African American	16	21		13	16		17	21		18	29		64	21	
American Indian	0	0		1	1		2	2		0	0		3	1	
Asian/Pacific Island	14	19		8	10		12	15		13	21		47	16	
Caucasian	37	49		50	61		40	49		24	38		151	50	
Hispanic/Latino	8	11		10	12		11	13		8	13		37	12	
Gender															
Male	37	49		42	51		42	51		29	46		150	50	
Female	38	51		40	49		40	49		34	54		152	50	
SES															
No	64	85		57	70		55	67		45	71		221	73	
Yes	11	15		25	30		27	33		18	29		81	27	
Limited English Proficiency															
No	65	87		62	76		62	76		57	90		246	81	
Yes	10	13		20	24		20	24		6	10		56	19	
Special Education															
No	75	100		80	98		77	94		61	97		293	97	
Yes	0	0		2	2		5	6		2	3		9	3	

## *Measures*

In this study, concurrent validity information was gathered on ISIP kindergarten, first, second, and third grade benchmark assessments compared to *Dynamic Indicators of Early Literacy Skills* kindergarten, first, second, and third grade benchmark assessments (DIBELS, Good, Kaminski, & Dill, 2002). Further, predictive validity information was gathered on ISIP compared to Texas Assessment of Knowledge and Skills (TAKS) at third grade.

**ISIP.** Subtests from ISIP included: Beginning Sound Fluency, Phonemic Blending Fluency, Letter Recognition Fluency, Letter Sound Fluency, Nonsense Word Fluency, Timed Reading with Meaning, and Comprehension.

In the *Beginning Sound Fluency* subtest, four items appear on the screen at once. The narrator says the name of each picture, as the box around it highlights. Then students are asked to click on the picture that has the same beginning sound as what the narrator says. The narrator then says one of the initial sounds. Students have 30 seconds to complete this activity. The activity timer is off during the warm-up, all instruction and during non-response re-tries. Both the total number of items correct and accuracy rate are taken into consideration in the student score. The score is then normalized to a fluency rate per minute. If the accuracy rate is in the range of chance (25% for this activity), a score of 0 is given.

In *Phonemic Blending Fluency* four items appear on the screen, with a box in the middle of the items that contains an animated side view of a head. The narrator says the name of each picture, as the box around it highlights. The narrator says one of the words phoneme by phoneme, as the animated head produces each sound. The child is asked to click on the picture showing a word that has been said phoneme by phoneme. Students have 30 seconds to complete this activity. The activity timer is off during the warm-up, all instruction and during non-response re-tries. The total number of items correct and accuracy rate are taken into consideration in the student score. The score is then normalized to a fluency rate per minute. If the accuracy rate is in the range of chance (25% for this activity), a score of 0 is given.

The *Letter Recognition Fluency* subtest asks students to identify the symbol for a letter's name. The computer randomly presents items representing various upper- and lowercase letters. Five letters appear on the screen for each item. The narrator asks the student to click on a particular letter. Students have 30 seconds to complete this activity. The activity timer is off during the warm-up, all instruction and during non-response re-tries. The total number of items correct and accuracy rate are taken into consideration in the student score. The score is then normalized to a fluency rate per minute. If the accuracy rate is in the range of chance (20% for this activity), a score of 0 is given.

In *Letter Sound Fluency* subtest uses the same game screen as *letter recognition*. But in this subtest, the narrator asks the student to click on the letter that makes a particular sound. Students have 30 seconds to complete this activity. The activity timer is off during the warm-up, all instruction and during non-response re-tries. The total number of items correct and accuracy rate are taken into consideration in the student score. The score is then normalized to a fluency

rate per minute. If the accuracy rate is in the range of chance (20% for this activity), a score of 0 is given.

The objective of the *Nonsense Word Fluency* subtest is for children to correctly identify non-words that are pronounced by the narrator. The narrator pronounces a non-word, and the children choose from four items that appear on the screen. Students have 30 seconds to complete this activity. The activity timer is off during the warm-up, all instruction and during non-response re-tries. The total number of items correct and accuracy rate are taken into consideration in the student score. The score is then normalized to a fluency rate per minute. If the accuracy rate is in the range of chance (25% for this activity), a score of 0 is given.

*Timed Reading with Meaning* is constructed in a very different manner than the other subtests. Children are assessed on their skills in reading grade level passages with meaning in a specified period of time. To assess text reading on the computer, a maze task is utilized in which every 5<sup>th</sup> to 8<sup>th</sup> word of grade-leveled passages is left blank from the text. The child is given 3 choices for each blank from which to choose the word that works in the sentence. This task has been shown to be highly correlated to measures of fluency and comprehension. Students have 2 minutes to engage in a story. The timer is off during the warm-up, during all instruction and during page turns. The total number of correct items, the number of words read to complete the correct items, and accuracy rate are taken into consideration in the student score. The score is then normalized to a fluency rate per minute. At moderate to low levels of accuracy on a high number of mazes, it is difficult to evaluate overall student performance. If a score cannot be calculated with a reasonable degree of confidence that it reflects the student's ability, a score of 0 is assigned. A student with an overall accuracy of less than 65% is assigned a score of 0.

In the *Comprehension* subtest, children are assessed on their ability to read and understand grade-leveled sentences. This is accomplished through matching sentences and pictures and sentence completion tasks. Matching sentences and pictures assesses students' knowledge of semantic and syntactic information where pictures can support their reading. Sentence completion measures the students' ability to use word meanings and word order to understand a sentence. Students have 2 minutes to read as many sentences as they can. Students are allowed up to 20 seconds for Grade 1 and 30 seconds for Grade 2 and 3 to complete an item. The additional time allowed for Grade 2 and 3 students is because of a greater number of words to be read on each item. Also, the test is discontinued if, after completing four items, the students' accuracy is less than 25%. The comprehension subtest score is derived from the number of words read on correctly answered questions, accuracy rate, and as an approximation of random chance answers. At moderate to low levels of accuracy on a high number of answered questions, it is difficult to evaluate overall student performance. If a score cannot be calculated with a reasonable degree of confidence that it reflects the student's ability, a score of 0 is assigned. For 1<sup>st</sup> and 2<sup>nd</sup> grade students with an overall accuracy of less than 60% and 3<sup>rd</sup> grade students with an overall accuracy of less than 70% a score of 0 is assigned. However, the student score is converted to a per minute score.

***DIBELS***. Individually administered subtests from *DIBELS* were administered to all participants and included Initial Sound Fluency, Letter Naming Fluency, Phoneme Segmentation Fluency, Nonsense Word Fluency, and Oral Reading Fluency.

*Initial Sound Fluency* (ISF) is a measure of phonological awareness that assesses a child's ability to recognize and produce the initial sound in an orally presented word. The examiner presents four pictures to the child, names each picture, and then asks the child to identify (i.e., point to or say) the picture that begins with the sound produced orally by the examiner. The score represents the number of initial sounds identified per minute.

*Letter Naming Fluency* (LNF) measures knowledge of letter names. Students are presented with a page of upper- and lowercase letters arranged in a random order and are asked to name as many letters as they can. The score represents the number of letters identified per minute.

*Phoneme Segmentation Fluency* (PSF) measures phonological awareness by assessing a student ability to segment three- and four-phoneme words into individual phonemes. The examiner orally presents words of three to four phonemes. Students are asked to verbally produce the individual phonemes for each word. The score represents the number of phonemes segmented correctly per minute.

*Nonsense Word Fluency* (NWF) measures letter-sound correspondence and the ability to blend letters into words in which letters represent their most common sounds. The student is presented an 8.5" x 11" sheet of paper with randomly ordered VC and CVC nonsense words (e.g., sig, rav, ov) and asked to verbally produce the individual letter sound of each letter or read the whole nonsense word. The score represents the number of phonemes read correctly per minute.

*Oral Reading Fluency* (ORF) passages are calibrated for the goal level of reading for each grade level. Student performance is measured by having students read a passage aloud for one minute. Words omitted, substituted, and hesitations of more than three seconds are scored as errors. Words self-corrected within three seconds are scored as accurate. The number of correct words per minute from the passage is the oral reading fluency rate.

### ***Data Collection Procedures***

Trained examiners with background clearance administered the DIBELS and ISIP assessments. DIBELS was administered individually. ISIP was administered in small groups of 1 examiner for no more than 5 children in the computer lab. To complete an ISIP assessment each child was seated in front of his/her own computer, logged on, and was administered ISIP. Each DIBELS required no more than 15 minutes per child to complete, depending on the grade of the child. Each ISIP session required no more than 25 minutes per group of children to complete, depending on the grade of the children.

### ***Examiner Training and Inter-rater Reliability***

All examiners were trained to administer ISIP to students in kindergarten through third grade by istation personnel. The same testers were trained to administer DIBELS. They were trained to administer DIBELS Initial Sound Fluency, Letter Naming Fluency, Phoneme Segmentation Fluency, Nonsense Word Fluency, and Oral Reading Fluency. This training was done by the project coordinator who was experienced at administering DIBELS and relied on the DIBELS manual and DIBELS Implementation CD-ROM. Fidelity reliability was checked by having testers all watch the same video segment of children taking DIBELS subtests, scoring the child, and discussing any disagreements in coding. Any tester whose scores were not within two points of the trainer's scores was given additional training until they could score within one or two points of the trainer's scores.

### ***Research Design***

Students were administered the ISIP subtests in four sessions to correspond to each of 4 benchmark assessment periods in which parallel forms of each subtest were administered. Under normal circumstances, students would be administered a beginning of year (BOY), 2 middle of year (MOY1 and MOY2), and 1 end of year (EOY) benchmark assessment administered about 2 months apart from each other. However, for the current study, students were administered all four benchmark assessments for each subtest within a 4-week window of time. To correct for a possible order effect, benchmark order was counterbalanced across students. One fourth of the sample was administered the benchmark assessments typically associated with the beginning of the year at time 1, another fourth took benchmark 1 at time 2, another fourth took it at time 3, and the remaining fourth took it at time 4. The process was repeated for the remaining benchmarks. Figure 1 delineates what subtests were administered at each grade for each benchmark. Figure 2 delineates how counterbalancing was enacted.

In the same 4-week window, students were also administered the three DIBELS benchmark assessments. DIBELS assessments were not counterbalanced, but were given in the order specified in the DIBLES manual. Students alternated between taking one DIBELS and one ISIP assessment (ISIP Time 1 → DIBELS Time 1 → ISIP Time 2 → DIBELS Time 2 → ISIP Time 3 → DIBELS Time 3 → ISIP Time 4). Students took no more than 2 ISIP assessments and 1 DIBELS assessment in any one week.

Figure 1. ISIP subtests administered at each grade for each benchmark.

Benchmark Assessments	K				1 <sup>st</sup> Grade				2 <sup>nd</sup> Grade				3 <sup>rd</sup> Grade			
	B	M	M	E	B	M	M	E	B	M	M	E	B	M	M	E
	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
		1	2			1	2			1	2			1	2	
<b>Beginning Sound Fluency (BSF)</b>	X	X	X													
<b>Letter Recognition Fluency (LRF)</b>	X	X	X	X	X											
<b>Letter Sound Fluency (LSF)</b>	X	X	X	X	X	X	X	X								
<b>Phoneme Blending Fluency (PBF)</b>			X	X	X	X	X	X								
<b>Nonsense Word Fluency (NWF)</b>				X	X	X	X	X	X	X	X	X				
<b>Comprehension (CMP)</b>							X	X	X	X	X	X	X	X	X	X
<b>Timed Reading with Meaning (TRM)</b>							X	X	X	X	X	X	X	X	X	X

Figure 2. Order of ISIP counterbalancing.

	<b>Time 1</b>	<b>Time 2</b>	<b>Time 3</b>	<b>Time 4</b>
<b>A order</b>	BOY	MOY1	MOY2	EOY
<b>B order</b>	MOY1	MOY2	EOY	BOY
<b>C order</b>	MOY2	EOY	BOY	MOY1
<b>D order</b>	EOY	BOY	MOY1	MOY2

## Results

### *ISIP Reliability*

The reliability of the ISIP benchmark assessments was examined by correlating each alternate form to every other alternate form for each subtest. Tables 3 - 20 present the results of these correlations using Pearson's  $r$ .

Table 3: Alternate form reliability for ISIP's Beginning Sound subtest for kindergarten.

	BOY	MOY1	MOY2
BOY	1.00	0.63	0.74
MOY1	0.63	1.00	0.68
MOY2	0.74	0.68	1.00

n = 75

Table 4: Alternate form reliability for ISIP's Phoneme Blending Fluency subtest for Kindergarten.

	MOY2	EOY
MOY2	1.00	0.81
EOY	0.81	1.00

n = 75

Table 5: Alternate form reliability for ISIP's Phonemic Blending Fluency for Grade 1.

	BOY	MOY1	MOY2	EOY
BOY	1.00	0.62	0.69	0.68
MOY1	0.62	1.00	0.61	0.64
MOY2	0.69	0.61	1.00	0.61
EOY	0.68	0.64	0.61	1.00

n = 82

Table 6: Alternate form reliability for ISIP's Phonemic Blending Fluency for Kindergarten and Grade I combined.

	BOY	MOY1	MOY2	EOY
BOY	1.00	0.62	0.69	0.68
MOY1	0.62	1.00	0.61	0.64
MOY2	0.69	0.61	1*	0.70*
EOY	0.68	0.64	0.70*	1*

n = 82

\*n = 157 kindergarten administered PBF at middle and end of year

Table 7: Alternate form reliability for ISIP's Letter Recognition subtest for Kindergarten.

	BOY	MOY1	MOY2	EOY
BOY	1.00	0.62	0.56	0.46
MOY1	0.62	1.00	0.62	0.64
MOY2	0.56	0.62	1.00	0.64
EOY	0.46	0.64	0.64	1.00

n = 75

Table 8: Alternate form reliability for ISIP's Letter Recognition Fluency for Kindergarten and Grade 1 combined.

	BOY*	MOY1	MOY2	EOY
BOY*	1.00	0.62	0.56	0.46
MOY1	0.62	1.00	0.62	0.64
MOY2	0.56	0.62	1.00	0.64
EOY	0.46	0.64	0.64	1.00

n = 75

\*n = 157

Table 9: Alternate form reliability for ISIP's Letter Sound subtest for Kindergarten.

	BOY	MOY1	MOY2	EOY
BOY	1.00	0.68	0.54	0.64
MOY1	0.68	1.00	0.56	0.68
MOY2	0.54	0.56	1.00	0.67
EOY	0.64	0.68	0.67	1.00

n = 75

Table 10: Alternate form reliability for ISIP's Letter Sound subtest for Grade 1.

	Boy	MOY1	MOY2	EOY
BOY	1.00	0.53	0.41	0.47
MOY1	0.53	1.00	0.54	0.58
MOY2	0.41	0.54	1.00	0.57
EOY	0.47	0.58	0.57	1.00

n = 82

Table 11: Alternate form reliability for ISIP's Letter Sound subtest for Kindergarten and Grade 1 combined.

	BOY	MOY1	MOY2	EOY
BOY	1.00	0.64	0.51	0.59
MOY1	0.64	1.00	0.57	0.66
MOY2	0.51	0.57	1.00	0.64
EOY	0.59	0.66	0.64	1.00

n = 157

Table 12: Alternate form reliability for ISIP's Nonsense Word subtest for Grade 1.

	BOY	MOY1	MOY2	EOY
BOY	1.00	0.90	0.77	0.89
MOY1	0.90	1.00	0.82	0.82
MOY2	0.77	0.82	1.00	0.84
EOY	0.89	0.82	0.84	1.00

Table 13: Alternate form reliability for ISIP's Nonsense Word subtest for Grade 2.

	BOY	MOY1	MOY2	EOY
BOY	1.00	0.75	0.62	0.52
MOY1	0.75	1.00	0.83	0.80
MOY2	0.62	0.83	1.00	0.73
EOY	0.52	0.80	0.73	1.00

n = 82

Table 14: Alternate form reliability for ISIP's Nonsense Word subtest for Grades 1 and 2 combined.

	BOY	MOY1	MOY2	EOY
BOY	1.00	0.85	0.73	0.73
MOY1	0.85	1.00	0.86	0.84
MOY2	0.73	0.86	1.00	0.81
EOY	0.73	0.84	0.81	1.00

n = 164

Table 15: Alternate form reliability for ISIP's Timed Reading with Meaning subtest for Grade 1.

	MOY2	EOY
MOY2	1.00	0.95
EOY	0.95	1.00

n = 82

Table 16: Alternate form reliability for ISIP's Timed Reading with Meaning subtest for Grade 2.

	BOY	MOY1	MOY2	EOY
BOY	1.00	0.82	0.82	0.87
MOY1	0.82	1.00	0.82	0.80
MOY2	0.82	0.82	1.00	0.84
EOY	0.87	0.80	0.84	1.00

n = 82

Table 17: Alternate form reliability for ISIP's Timed Reading with Meaning subtest for Grade 3.

	BOY	MOY1	MOY2	EOY
BOY	1.00	0.81	0.76	0.64
MOY1	0.81	1.00	0.88	0.78
MOY2	0.76	0.88	1.00	0.90
EOY	0.64	0.78	0.90	1.00

n = 63

Table 18: Alternate form reliability for ISIP's Timed Reading with Meaning subtest for Grades 1-3 combined.

	BOY	MOY1	MOY2	EOY
BOY	1*	0.85	0.81	0.77
MOY1	0.85	1.00	0.86	0.80
MOY2	0.81	0.86	1*	0.92*
EOY	0.77	0.90	0.92*	1*

n = 145  
\*n = 227

Table 19: Alternate form reliability for ISIP's Comprehension for Grade 1.

	MOY2	EOY
MOY2	1.00	0.89
EOY	0.89	1.00

n = 82

Table 20: Alternate form reliability for ISIP's Comprehension for Grade 2.

	BOY	MOY1	MOY2	EOY
BOY	1.00	0.86	0.84	0.86
MOY1	0.86	1.00	1.00	0.85
MOY2	0.84	1.00	1.00	0.81
EOY	0.86	0.85	0.81	1.00

n = 82

Table 21: Alternate form reliability for ISIP's Comprehension for Grade 3.

	BOY	MOY1	MOY2	EOY
BOY	1.00	0.77	0.68	0.73
MOY1	0.77	1.00	0.66	0.87
MOY2	0.68	0.66	1.00	0.64
EOY	0.73	0.87	0.64	1.00

n = 63

Table 22: Alternate form reliability for ISIP's Comprehension for Grades 1- 3 combined.

	BOY	MOY1	MOY2	EOY
BOY	1.00	0.85	0.81	0.84
MOY1	0.85	1.00	0.83	0.88
MOY2	0.81	0.83	1*	0.89*
EOY	0.84	0.88	0.89*	1*

n = 145

\*n = 227

**ISIP Clusters.** We also created combined scores on subtests that were related to the same construct to form 4 clusters:

1. Phonemic Awareness (Beginning Sound + Phoneme Blending),
2. Letter Knowledge (Letter Recognition + Letter Sound Fluency),
3. Alphabetic Knowledge (Letter Sound Fluency + Nonsense Word Fluency),
4. Reading with Meaning (Timed Reading with Meaning + Comprehension).

The reliability of these ISIP clusters was examined by correlating each alternate cluster form to every other alternate form for each cluster. Tables 23 - 29 present the results of these correlations using Pearson's *r*.

Table 23: Alternate form reliability for ISIP's Phonemic Awareness Cluster for Kindergarten and Grade 1 combined.

	BOY*	MOY1	MOY2	EOY
BOY*	1.00	0.82	0.76	0.76
MOY1	0.82	1.00	0.85	0.87
MOY2	0.76	0.85	1.00	0.86
EOY	0.76	0.87	0.86	1.00

n=75

\*n =157

Table 24: Alternate form reliability for ISIP's Letter Knowledge Cluster for Kindergarten and Grade 1 combined.

	BOY*	MOY1	MOY2	EOY
BOY*	1.00*	0.72	0.66	0.664
MOY1	0.72	1	0.75	0.769
MOY2	0.66	0.75	1	0.763
EOY	0.66	0.77	0.76	1

n = 75  
\*n = 157

Table 25: Table 25: Alternate form reliability for ISIP's Alphabetic Knowledge Cluster for Kindergarten and Grade 1.

	BOY	MOY1	MOY2	EOY*
BOY	1.00	0.90	0.75	1.00
MOY1	0.90	1.00	0.83	0.84
MOY2	0.75	0.83	1.00	0.85
EOY*	0.90	0.84	0.85	1.00

n = 82  
\*n = 157

Table 26: Alternate form reliability for ISIP's Reading with Meaning Cluster for Grade 1.

	MOY2	EOY
MOY2	1	0.95
EOY	0.95	1

n = 82

Table 27: Alternate form reliability for ISIP's Reading with Meaning Cluster for Grade 2

	BOY	MOY1	MOY2	EOY
BOY	1.00	0.95	0.94	0.94
MOY1	0.95	1.00	0.96	0.90
MOY2	0.94	0.96	1.00	0.92
MOY3	0.94	0.90	0.92	1.00

n = 82

Table 28: Alternate form reliability for ISIP's Reading with Meaning Cluster for Grade 3

	BOY	MOY1	MOY2	EOY
BOY	1.00	0.91	0.83	0.86
MOY1	0.91	1.00	0.89	0.91
MOY2	0.83	0.89	1.00	0.91
WOY	0.86	0.91	0.91	1.00

n=63

Table 29: Alternate form reliability for ISIP's Reading with Meaning Cluster for Grades 1-3 combined.

	BOY	MOY1	MOY2	EOY
BOY	1.00	0.95	0.92	0.92
MOY1	0.95	1.00	0.95	0.92
MOY2	0.92	0.95	1.00*	0.99*
EOY	0.92	0.92	0.99*	1.00*

n = 145

\*n = 227

### ***DIBELS Reliability***

The reliability of the DIBELS benchmark assessments was examined by correlating each alternate form to every other alternate form for individual subtests. Tables 30 – 41 present the results of these correlations using Pearson's  $r$ .

Table 30: Alternate form reliability for DIBELS' Initial Sound Fluency for Kindergarten.

	BOY	MOY
BOY	1	0.42
MOY	0.42	1

n = 75

Table 31: Alternate form reliability for DIBELS' Letter Naming Fluency for Kindergarten.

	BOY	MOY	EOY
BOY	1	0.92	0.87
MOY	0.92	1	0.9
EOY	0.87	0.9	1

n = 75

Table 32: Alternate form reliability for DIBELS' Phoneme Segmentation Fluency for Kindergarten.

	MOY	EOY
MOY	1	0.54
EOY	0.54	1

n = 75

Table 33: Alternate form reliability for DIBELS' Phoneme Segmentation Fluency for Grade 1.

	BOY	MOY	EOY
BOY	1	0.53	0.54
MOY	0.53	1	0.59
EOY	0.54	0.59	1

n= 82

Table 34: Alternate form reliability for DIBELS' Phoneme Segmentation Fluency for Kindergarten and Grade 1 combined.

	BOY	MOY	EOY
BOY	1.00	0.53	0.54
MOY	0.53	1.00*	0.58*
EOY	0.54	0.58*	1.00*

\*n=157 kindergarten administered PSF at middle and end of year

Table 35: Alternate form reliability for DIBELS' Nonsense Word Fluency for Kindergarten.

	MOY	EOY
MOY	1	0.915
EOY	0.915	1

n =75

Table 36: Alternate form reliability for DIBELS' Nonsense Word Fluency for Grade 1.

	BOY	MOY	EOY
BOY	1	0.81	0.85
MOY	0.81	1	0.82
EOY	0.85	0.82	1

n= 82

Table 37: Alternate form reliability for DIBELS' Nonsense Word Fluency for Kindergarten - Grade 2 combined.

	BOY	MOY	EOY
BOY	1.00*	.81**	0.85*
MOY	.81**	1.00	0.91
EOY	0.85*	0.91	1.00

n=157 kindergarten and grade 1

\*n=164 grade 1 and grade 2

\*\*n=82 grade 1 only

Table 38: Alternate form reliability for DIBELS' Oral Reading Fluency for Grade 1.

	MOY	EOY
MOY	1	0.9
EOY	0.9	1

n = 82

Table 39: Alternate form reliability for DIBELS' Oral Reading Fluency for Grade 2.

	BOY	MOY	EOY
BOY	1	0.95	0.94
MOY	0.95	1	0.94
EOY	0.94	0.94	1

n = 82

Table 40: Alternate form reliability for DIBELS' Oral Reading Fluency for Grade 3.

	BOY	MOY	EOY
BOY	1	0.93	0.92
MOY	0.93	1	0.92
EOY	0.92	0.92	1

n = 63

Table 41: Alternate form reliability for DIBELS' Oral Reading Fluency Grades 1 - 3 combined.

	BOY	MOY	EOY
BOY	1.00	0.94	0.93
MOY	0.94	1.00*	0.94*
EOY	0.93	0.94*	1.00*

n = 145

\*n = 227 first grade administered ORF at middle and end of year

### *Concurrent Validity of ISIP Benchmark Assessments*

To establish concurrent validity of ISIP Benchmark assessments, we correlated the ISIP subtests to DIBELS subtests in Kindergarten through Grade 3. Because of the restricted range of our sample, we applied the following correction for restriction of range,

$$r_u = \frac{rr * (SD_u / SD_r)}{\sqrt{1 - rr^2 + rr^2 * (SD_u^2 / SD_r^2)}}$$

In which the Observed Restricted Correlation is represented as  $rr$ , the Standard Deviation of the restricted sample is represented as  $SD_r$ . The standard deviation of the unrestricted sample (i.e., the norm sample) is represented as  $SD_u$ , and the estimated unrestricted correlation is represented as  $r_u$  (Guilford & Fruchter 1978). This formula has as the top value the result of multiplying the restricted correlation times the ratio of the unrestricted standard deviation (SD) divided by the restricted SD. The bottom value is a square root of the following calculation: Take unity (1.0) and subtract the squared restricted correlation. Add the quantity of the squared restricted correlation times the ratio of the two variances (squared SDs) — the unrestricted variance divided by the restricted variance.

Correlations matrixes for individual ISIP subtests compared to DIBELS subtests are presented in Tables 42-47.

Table 42: Correlations of ISIP Beginning Sound Fluency Benchmarks and DIBELS Initial Sound Benchmarks for Kindergarten.

ISIP	DIBELS	
	BOY	MOY
BOY	0.65	0.82
MOY1	0.88	0.93
MOY2	0.66	0.89

N=75

Table 43: Correlations of ISIP Phoneme Blending Fluency and DIBELS Phoneme Segmentation Fluency Benchmarks for Kindergarten and Grade 1.

	ISIP Phoneme Blending Fluency	DIBELS Phoneme Segmentation Fluency		
		BOY	MOY	EOY
		Grade k (n = 75)	MOY2	
	EOY		0.55	0.48
Grade 1 (n = 82)	BOY	0.06	0.01	0.17
	MOY1	0.03	0.00	0.03
	MOY2	0.05	0.04	0.11
	EOY	0.04	0.01	0.18

Table 44: Correlations of ISIP Letter Recognition Benchmarks and DIBELS Letter Name Fluency Benchmarks for Kindergarten and Grade 1.

	ISIP Letter Recognition Fluency	DIBELS Letter Naming Fluency		
		BOY	MOY	EOY
		Grade K (n=75)	BOY	0.35
	MOY1	0.55	0.54	0.41
	MOY2	0.32	0.32	0.23
	EOY	0.49	0.53	0.44
Grade 1 (n=82)	BOY	0.36		

Table 45: Correlations of ISIP Letter Sound Fluency Benchmarks and DIBELS Letter Name Fluency Benchmarks for Kindergarten and Grade 1.

	ISIP Letter Sound Fluency	DIBELS Letter Name Fluency		
		BOY	MOY	EOY
Kindergarten (n=75)	BOY	0.47	0.47	0.34
	MOY1	0.58	0.60	0.47
	MOY2	0.49	0.45	0.45
	EOY	0.49	0.48	0.34
Grade 1 (n=82)	BOY	0.18		
	MOY1	0.31		
	MOY2	0.30		
	EOY	0.24		

Table 46: Correlations of ISIP Nonsense Word Fluency Benchmarks and DIBELS Nonsense Word Fluency Benchmarks for Grades K, 1, and 2.

	ISIP Nonsense Word Fluency	DIBELS Nonsense Word Fluency		
		BOY	MOY	EOY
Grade K (n = 75)	EOY		0.82	0.83
	BOY	0.79	0.78	0.77
Grade 1 (n = 82)	MOY1	0.77	0.90	0.77
	MOY2	0.66	0.76	0.67
	EOY	0.86	0.90	0.85
Grade 2 (n = 82)	BOY	0.66		
	MOY1	0.68		
	MOY2	0.67		
	EOY	0.66		

Table 47: Correlations of ISIP Timed Reading with Meaning Benchmarks and DIBELS Oral Reading Fluency Benchmarks for Grades 1 -3.

ISIP Timed Reading with Meaning		DIBELS Oral Reading Fluency		
		BOY	MOY	EOY
Grade 1 (n = 75)	MOY2		0.88	0.95
	EOY		0.90	0.94
Grade 2 (n=82)	BOY	0.85	0.87	0.89
	MOY1	0.88	0.82	0.81
	MOY2	0.86	0.91	0.92
	EOY	0.86	0.89	0.91
Grade 3 (n = 63)	BOY	0.93	0.95	0.94
	MOY1	0.96	0.97	0.97
	MOY2	0.94	0.95	0.97
	EOY	0.79	0.81	0.79

Table 48: Correlations of ISIP Comprehension Benchmarks and DIBELS Oral Reading Fluency Benchmarks for Grades 1 -3.

ISIP Comprehension		DIBELS Oral Reading Fluency		
		BOY	MOY	EOY
Grade 1 (n = 82)	MOY 2		0.74	0.85
	EOY		0.76	0.86
Grade 2 (n = 82)	BOY	0.79	0.84	0.82
	MOY1	0.80	0.84	0.82
	MOY2	0.83	0.86	0.83
	EOY	0.76	0.77	0.75
Grade 3 (n = 63)	BOY	0.72	0.75	0.71
	MOY1	0.76	0.77	0.74
	MOY2	0.74	0.76	0.76
	EOY	0.71	0.75	0.73

**Clusters.** Correlations matrixes for ISIP clusters compared to DIBELS subtests are presented in Tables 49-52.

Table 49: Correlations of ISIP Phonemic Awareness Cluster and DIBELS Initial Sound Fluency Benchmarks for Kindergarten.

ISIP Phonemic Awareness Cluster	DIBELS Initial Sound Fluency	
	BOY	MOY
MOY2	0.48	0.54

n = 75

Table 50: Correlations of ISIP Phonemic Awareness Cluster and DIBELS Phoneme Segmentation Fluency Benchmarks for Kindergarten.

ISIP Phonemic Awareness	DIBELS Phoneme Segmentation Fluency	
	MOY	EOY
MOY2	0.56	0.46

n = 75

Table 51: Correlations of ISIP Letter Knowledge Cluster and DIBELS Letter Naming Fluency Benchmarks for Grade K and 1.

	ISIP Letter Knowledge	DIBELS Letter Naming Fluency		
		BOY	MOY	EOY
Kindergarten n = 75	BOY	0.56	0.56	0.53
	MOY1	0.72	0.72	0.58
	MOY2	0.58	0.55	0.51
	EOY	0.62	0.64	0.52
Grade 1 n = 82	BOY	0.49		

Table 52: Correlations of ISIP Alphabetic Skills Cluster and DIBELS Nonsense Word Fluency Benchmarks for Grades K and 1.

	ISIP Alphabetic Skills	DIBELS Nonsense Word Fluency		
		BOY	MOY	EOY
Grade K (n = 82)	EOY		0.72	0.73
	BOY	0.78	0.67	0.66
Grade 1 (n = 82)	MOY1	0.67	0.71	0.68
	MOY2	0.57	0.66	0.58
	EOY	0.76	0.90	0.75

Table 52: Correlations of ISIP Reading with Meaning Cluster and DIBELS Oral Reading Fluency Benchmarks for Grades 1, 2, and 3.

	ISIP Reading with Meaning Cluster	DIBELS Oral Reading Fluency		
		BOY	MOY	EOY
Grade 1 (n = 62)	MOY2		0.91	0.91
	EOY		0.87	0.91
Grade 2 (n = 82)	BOY	0.90	0.93	0.93
	MOY1	0.89	0.93	0.91
	MOY2	0.84	0.88	0.95
	EOY	0.91	0.92	0.93
Grade 3 (n = 63)	BOY	0.94	0.95	0.95
	MOY1	0.86	0.88	0.86
	MOY2	0.94	0.95	0.88
	EOY	0.87	1.00	0.88

### *Identification of Risk Status*

Both ISIP and DIBELS identify risk status based on performance on each subtest. Both tests classify students into 3 categories (Low-risk, Some-Risk, and High-Risk). To determine if the same students were being identified into these 3 risk categories on each test, the number of agreements of classifications combining grades were cross tabulated and Cohen's Kappa computed to determine the independence of each rankings. Thus, low p values correspond to

better agreement. Tables 53 -58 present identification of risk status using ISIP criteria compared to identification of risk status using DIBELS criteria.

Table 53: Crosstabulation of Risk Status for ISIP Beginning Sound Fluency and DIBELS Initial Sound Fluency.

		ISIP Beginning Sound Fluency		
		High Risk	Some Risk	Low Risk
DIBELS Initial Sound Fluency	High Risk	<b>0</b>	0	1
	Some Risk	1	<b>4</b>	15
	Low Risk	2	8	<b>44</b>

Kappa = .06, p = .59

Table 54: Crosstabulation of Risk Status for ISIP Letter Recognition Fluency and DIBELS Letter Naming Fluency.

		ISIP Letter Recognition Fluency		
		High Risk	Some Risk	Low Risk
DIBELS Letter Naming Fluency	High Risk	<b>6</b>	4	3
	Some Risk	3	<b>3</b>	7
	Low Risk	9	29	<b>93</b>

Kappa = .27, p = .001

Table 55: Crosstabulation of Risk Status for ISIP Letter Sound Fluency and DIBELS Letter Naming Fluency.

		ISIP Letter Sound Fluency		
		High Risk	Some Risk	Low Risk
DIBELS Letter Naming Fluency	High Risk	<b>6</b>	3	4
	Some Risk	6	<b>4</b>	3
	Low Risk	23	41	<b>67</b>

Kappa = .10, p = .03

Table 56: Crosstabulation of Risk Status for ISIP Phoneme Blending Fluency and DIBELS Phoneme Segmentation Fluency.

		ISIP Phoneme Blending Fluency		
		High Risk	Some Risk	Low Risk
DIBELS Phoneme Segmentation Fluency	High Risk	<b>1</b>	0	0
	Some Risk	1	<b>0</b>	6
	Low Risk	1	3	<b>145</b>

Kappa = .39, p = .00

Table 57: Crosstabulation of Risk Status for ISIP Nonsense Word Fluency and DIBELS Nonsense Word Fluency.

		ISIP Nonsense Word Fluency		
		High Risk	Some Risk	Low Risk
DIBELS Nonsense Word Fluency	High Risk	<b>11</b>	0	9
	Some Risk	4	<b>5</b>	10
	Low Risk	4	11	<b>195</b>

Kappa = .61, p = .00

Table 58: Crosstabulation of Risk Status for ISIP Timed Reading with Meaning and DIBELS Oral Reading Fluency.

		ISIP Timed Reading with Meaning		
		High Risk	Some Risk	Low Risk
DIBELS Oral Reading Fluency	High Risk	<b>28</b>	4	4
	Some Risk	7	<b>18</b>	28
	Low Risk	8	9	<b>121</b>

Kappa = .54, p = .00

### *Predictive Validity*

In addition to comparing ISIP and DIBELS subtests administered in the same window of time, we determined how well ISIP and DIBELS measures given to students in Grade 3 predicted their performance on the TAKS reading test administered in the school year. The same correction for restricted range was applied to these data.

Table 59: Correlation of ISIP and DIBELS Grade 3 Benchmarks to the Third-Grade Reading TAKS scaled scores.

	ISIP			DIBELS	
	Comprehension	Timed Reading with Meaning	Reading with Meaning Cluster		Oral Reading Fluency
BOY	0.57	0.45	0.54	BOY	0.46
MOY1	0.49	0.56	0.56	MOY	0.45
MOY2	0.67	0.53	0.64		
EOY	0.53	0.53	0.58	EOY	0.37
Average of Benchmarks	0.65	0.57	0.62		0.44

n = 60

## Discussion

The hypothesis for this study was that ISIP Benchmarks will be as valid an indicator of reading skills, and will identify students who are “at risk” of reading failure as well as a well-accepted, comparable assessment that is teacher administered (i.e., DIBELS), and that at 3<sup>rd</sup> grade, ISIP will be as predictive of 3<sup>rd</sup>-grade TAKS reading performance as DIBELS. In this section, we discuss our results in terms of our 4 original research questions.

### Question 1

Our first question was concerned with the reliability of each ISIP alternate form for each Benchmark and for each DIBELS alternate form for each Benchmark and if the alternate form reliabilities of these two tests were in line with one another.

**Alternate Form ISIP Correlations:** A majority of ISIP subtests demonstrate adequate alternate form reliability at most grades, although for several subtests, the BOY administration appeared to be less reliable than other administrations. We attribute this to a learning affect associated with taking a particular task for the first time. Our plan is to add additional practice items prior to administering BOY subtests.

We ascribed highly reliable status to subtests with Pearson’s  $r$  coefficient at or above .80 on all administrations, allowing administrations at BOY to fall in to the .70 - .79 range. ISIP subtests that demonstrate high levels of alternate form reliability include:

- Phoneme Blending Fluency at K ( $r = .81$ )
- Nonsense Word Fluency at 1<sup>st</sup> Grade ( $r = .77$  (BOY to MOY2) - .90)
- Nonsense Word Fluency at Grade 1,2 ( $r = .73$  (BOY to MOY2 and EOY) - .85)

- Timed Reading with Meaning at 1<sup>st</sup> Grade ( $r = .95$ )
- Timed Reading with Meaning at 2<sup>nd</sup>-Grade ( $r = .82 - .87$ )
- Timed Reading with Meaning at 3<sup>rd</sup>-Grade ( $r = .64$  (BOY to EOY) - .87)
- Timed Reading with Meaning across Grades 1-3 ( $r = .77$  (BOY to EOY) - .92)
- Comprehension at 1<sup>st</sup>-Grade ( $r = .89$ )
- Comprehension at 2<sup>nd</sup>-Grade ( $r = .81 - .86$ )
- Comprehension across Grades 1-3 ( $r = .81 - .89$ )
- Phonemic Awareness Cluster at Grades K-1 ( $r = .76$  (BOY to EOY) - .87)
- Alphabetic Knowledge Cluster at Grades K-1 (.75 (BOY to EOY) - .90)
- Reading with Meaning Cluster at 1<sup>st</sup>-Grade ( $r = .95$ )
- Reading with Meaning Cluster at 2<sup>nd</sup>-Grade ( $r = .90 - .95$ )
- Reading with Meaning Cluster at 3<sup>rd</sup>-Grade ( $r = .83 - .91$ )
- Reading with Meaning Cluster at Grades 1-3 ( $r = .92 - .99$ )

We ascribed moderately reliable status to subtests with Pearson's  $r$  coefficient between .69 and .79 on most administrations. In some cases, particular administration of a subtest were above  $r = .80$ . ISIP subtests that demonstrate moderate levels of alternate form reliability include:

- Beginning Sound Fluency at K ( $r = .63 - .74$ )
- Phoneme Blending Fluency at 1<sup>st</sup>-Grade ( $r = .62 - .69$ )
- Phoneme Blending Fluency at Grades K, 1 ( $r = .61 - .70$ )
- Comprehension at 3<sup>rd</sup> Grade ( $r = .66 - .87$ )
- Letter Knowledge Cluster at Grade k, 1 ( $r = .66 - .77$ )

We ascribed low reliability status to subtests with Pearson's  $r$  coefficient below .60 on any administration. ISIP subtests that demonstrate low levels of alternate form reliability include:

- Letter Recognition Fluency at K ( $r = .46 - .64$ )
- Letter Sound Fluency at K ( $r = .54 - .68$ )
- Letter Sound Fluency at K ( $r = .54 - .68$ )
- Letter Sound Fluency at K ( $r = .54 - .68$ )
- Nonsense Word Fluency at 2<sup>nd</sup>-Grade ( $r = .52$  (BOY to EOY) - .83)

**Alternate Form DIBELS Correlations.** The majority of DIBELS subtests also demonstrated high alternate form reliability with the same sample of children.

We ascribed highly reliable status to subtests with Pearson's  $r$  coefficient at or above .80 on all administrations, allowing administrations at BOY to fall in to the .70 - .79 range. DIBELS subtests that demonstrate high levels of alternate form reliability include:

- Letter Naming Fluency at K ( $r = .87 - .92$ )
- Nonsense Word Fluency at K ( $r = .92$ )
- Nonsense Word Fluency at 1<sup>st</sup>-Grade ( $r = .81 - .85$ )
- Nonsense Word Fluency at Grades K, 1 ( $r = .81 - .91$ )
- Oral Reading Fluency at 1<sup>st</sup>-Grade ( $r = .90$ )

- Oral Reading Fluency at 2<sup>nd</sup>-Grade ( $r = .94 - .95$ )
- Oral Reading Fluency at 3<sup>rd</sup>-Grade ( $r = .92 - .93$ )
- Oral Reading Fluency at Grades 1-3 ( $r = .93 - .94$ )

We ascribed moderately reliable status to subtests with Pearson's  $r$  coefficient between .69 and .79 on most administrations. In some cases, particular administration of a subtest were above  $r = .80$ . DIBELS subtests that demonstrate moderate levels of alternate form reliability include:

- Phoneme Segmentation Fluency at K ( $r = .54$ )
- Phoneme Segmentation Fluency at 2<sup>nd</sup>-Grade ( $r = .53 - .59$ )
- Phoneme Segmentation Fluency at Grades K & 1 ( $r = .53 - .58$ )

We ascribed low reliability status to any subtest with Pearson's  $r$  coefficient below .60 on any administration. One DIBELS subtest demonstrated low levels of alternate form reliability:

- Initial Sound Fluency at K ( $r = .42$ )

In general, it appears that both ISIP and DIBELS have similar alternate form reliabilities across benchmarks. However, given that several subtests on both tests do not achieve high levels of alternate form reliability, it should be expected that concurrent validity coefficients comparing these same subtests will be somewhat low.

### **Question 2:**

For question 2, we wanted to know the concurrent validity coefficients between ISIP and DIBELS for subtest measuring the same constructs. Many ISIP and DIBELS subtests demonstrated high levels concurrent validity as reflected in Pearson's attenuated  $r$  coefficient corrected for restricted range. The following ISIP and DIBELS subtests were found to be highly correlated on most administrations, with the exception of BOY administrations. In some cases, comparisons using BOY were in the moderate range:

- ISIP Beginning Sound Fluency & DIBELS Initial Sound Fluency ( $r = .65 - .93$ )
- ISIP Nonsense Word Fluency & DIBELS Nonsense Word Fluency at K ( $r = .82 - .83$ )
- ISIP Nonsense Word Fluency & DIBELS Nonsense Word Fluency at 1<sup>st</sup>-Grade ( $r = .66 - .98$ )
- ISIP Timed Reading with Meaning & DIBELS Oral Reading Fluency at 1<sup>st</sup>-Grade ( $r = .88 - .95$ )
- ISIP Timed Reading with Meaning & DIBELS Oral Reading Fluency at 2<sup>nd</sup>-Grade ( $r = .81 - .92$ )
- ISIP Timed Reading with Meaning & DIBELS Oral Reading Fluency at 3<sup>rd</sup>-Grade ( $r = .79 - .97$ )
- ISIP Comprehension & DIBELS Oral Reading Fluency at 1<sup>st</sup>-Grade ( $r = .74 - .86$ )
- ISIP Comprehension & DIBELS Oral Reading Fluency at 2<sup>nd</sup>-Grade ( $r = .75 - .86$ )

- ISIP Reading with Meaning Cluster and DIBELS Oral Reading Fluency at 1<sup>st</sup>-Grade ( $r = .87-.91$ )
- ISIP Comprehension & DIBELS Oral Reading Fluency at 1<sup>st</sup>-Grade ( $r = .74 - .86$ )
- ISIP Reading with Meaning Cluster and DIBELS Oral Reading Fluency at 3<sup>rd</sup>-Grade ( $r = .87 - .95$ )

The relationship between these sets of subtests demonstrates that they are both tapping into essentially the same construct. It is also logical that, while high, ISIP comprehension and DIBELS Oral Reading fluency have somewhat lower correlations than Timed Reading with Meaning and DIBELS Oral Reading Fluency, since Timed Reading with Meaning was designed specifically to test fluency, while comprehension was not. Even so, the relationship between comprehension and oral reading fluency is well documented, especially in the lower grades. This decreasing relationship is apparent in our data, with the relationship between ISIP comprehension and DIBELS Oral Reading Fluency showing high correlation at 1<sup>st</sup> and 2<sup>nd</sup> grade, but only moderate correlation at 3<sup>rd</sup> grade (see below).

The following ISIP and DIBELS subtests were found to be correlated at the moderate level:

- ISIP Beginning Phoneme Fluency & DIBELS Phoneme Segmentation Fluency at Grade K ( $r = .40 - .55$ )
- ISIP Letter Sound Fluency & DIBELS Letter Naming Fluency at Grade K ( $r = .34 - .60$ )
- ISIP Nonsense Word Fluency & DIBELS Nonsense Word Fluency at 2<sup>nd</sup>-Grade ( $r = .66 - .68$ )
- ISIP Comprehension & DIBELS Oral Reading Fluency at 3<sup>rd</sup>-Grade ( $r = .71 - .77$ )
- ISIP Phonemic Awareness Cluster & DIBELS Initial Sound Fluency at Grade K ( $r = .48 - .54$ )
- ISIP Alphabetic Knowledge Cluster & DIBELS Phoneme Segmentation Fluency at Grade K ( $r = .46 - .56$ )
- ISIP Alphabetic Knowledge Cluster & DIBELS Nonsense Word Fluency at Grade K ( $r = .72 - .73$ )
- ISIP Alphabetic Knowledge Cluster & DIBELS Nonsense Word Fluency at 1<sup>st</sup>-Grade ( $r = .57 - .90$ )
- ISIP Letter Knowledge Cluster & DIBELS Letter Naming Fluency at Grade K ( $r = .51 - .72$ )
- ISIP Letter Knowledge Cluster & DIBELS Letter Naming Fluency at 1<sup>st</sup>-Grade ( $r = .49$ )

The following ISIP and DIBELS subtests were found to be correlated at the weak level:

- ISIP Beginning Phoneme Fluency & DIBELS Phoneme Segmentation Fluency at 1<sup>st</sup>-Grade ( $r = .00 - .18$ )
- ISIP Letter Recognition Fluency and DIBELS Letter Naming Fluency at Grade K ( $r = .23 - .55$ )

- ISIP Letter Recognition Fluency and DIBELS Letter Naming Fluency at 1<sup>st</sup>-Grade ( $r = .36$ )
- ISIP Letter Sound Fluency and DIBELS Letter Naming Fluency at 1<sup>st</sup>-Grade ( $r = .18 - .31$ )

The moderate to low correlations observed between ISIP and DIBELS subtests are logical in that, with the exception of ISIP and DIBELS Nonsense Word Fluency, none of these subtest comparisons are designed to test exactly the same aspect of the same construct. For example, ISIP Beginning Sound Fluency is more closely related in design to DIBELS Initial Sound Fluency than to Phoneme Segmentation, even though both DIBELS subtests assess aspects of Phonemic Awareness. Thus, the fact that ISIP correlates highly to DIBELS Initial Sound Fluency and only moderately with Phoneme Segmentation is to be expected. Similarly, ISIP Letter Sound Fluency and the Alphabetic Knowledge Cluster are designed to assess if students recognize the sounds and or names of specific letters. However, DIBELS Letter Naming Fluency is really a test of Rapid Automatic Naming and is designed to test a child's lexical retrieval ability, which goes beyond simple letter knowledge, thus, lower correlations between ISIP letter naming or Letter Recognition tasks and DIBELS Letter Naming Fluency are expected.

### ***Question 3***

For Question 3, we asked if ISIP and DIBELS identify the same children as high risk, some risk, and low risk? As would be expected, those subtests that correlate less well to one another also identify children into the three risk categories less consistently. Conversely, the better related subtests are, the more consistently they identify the same children into the same categories. However, given that we have already demonstrated that those subtests that are correlated less well to one another also are less related in terms of their underlying construct, it is to be expected that less correlated subtests would identify different children in the various risk categories, while more correlated tests would be more likely to identify the same children into the same risk categories. The only exception to this Observation is the assignment of risk status between ISIP Phoneme Blending and DIBELS Phoneme Segmentation Fluency. While these subtests were weakly correlated, each subtest demonstrated moderate agreement in terms of risk status identification. The highest level of risk status agreements was observed between ISIP and DIBELS Nonsense Word Fluency subtests and between ISIP Timed Reading with Meaning and DIBELS Oral Reading Fluency.

### ***Question 4***

For Question 4, we wanted to know the predictive validity of both ISIP and DIBELS measures given at 3<sup>rd</sup>-Grade when compared to the Texas 3<sup>rd</sup>-Grade TAKS in reading? As is apparent in Table 59, both ISIP and DIBELS 3<sup>rd</sup>-Grade measures correlate moderately with performance on the 3<sup>rd</sup>-Grade Reading TAKS among students in our sample. However, it also appears that ISIP measures correlate somewhat more strongly to 3<sup>rd</sup>-Grade TAKS reading performance than DIBELS. It is also worth noting that ISIP Timed Reading with Meaning and Comprehension measures and DIBELS Oral Reading Fluency were less correlated to each other at 3<sup>rd</sup>-Grade than at 1<sup>st</sup> and 2<sup>nd</sup>-Grades. However, when compared to TAKS, it appears that both ISIP measures are actually somewhat more predictive of TAKS reading than DIBELS Oral

Reading Fluency. While we can only speculate as to why this might be so, it seems logical to assume that because both ISIP measures tap into comprehension and because 3<sup>rd</sup>-Grade TAKS is largely a comprehension measure, ISIP and TAKS might be better aligned than DIBELS, which is more a measure of fluency. While fluency is certainly a dimension of comprehension, it is also true that TAKS is an untimed test, and thus, the influence of fluency may be less related to TAKS performance than other comprehension measures that do include time as a dimension. Credibility for this line of reasoning is supported by the fact that the ISIP Comprehension subtest predicts TAKS performance more strongly than either ISIP Timed Reading with Meaning or DIBELS Oral Reading Fluency. However, Timed Reading with meaning, which does require that children monitor comprehension as they read connected text, predicts TAKS somewhat better than DIBELS Oral Reading Fluency.