Comparing the *Observation Survey* to DIBELS

Francisco X. Gómez-Bellengé

The Ohio State University

2008 National Reading Recovery Conference, Columbus, OH

12 February 2008

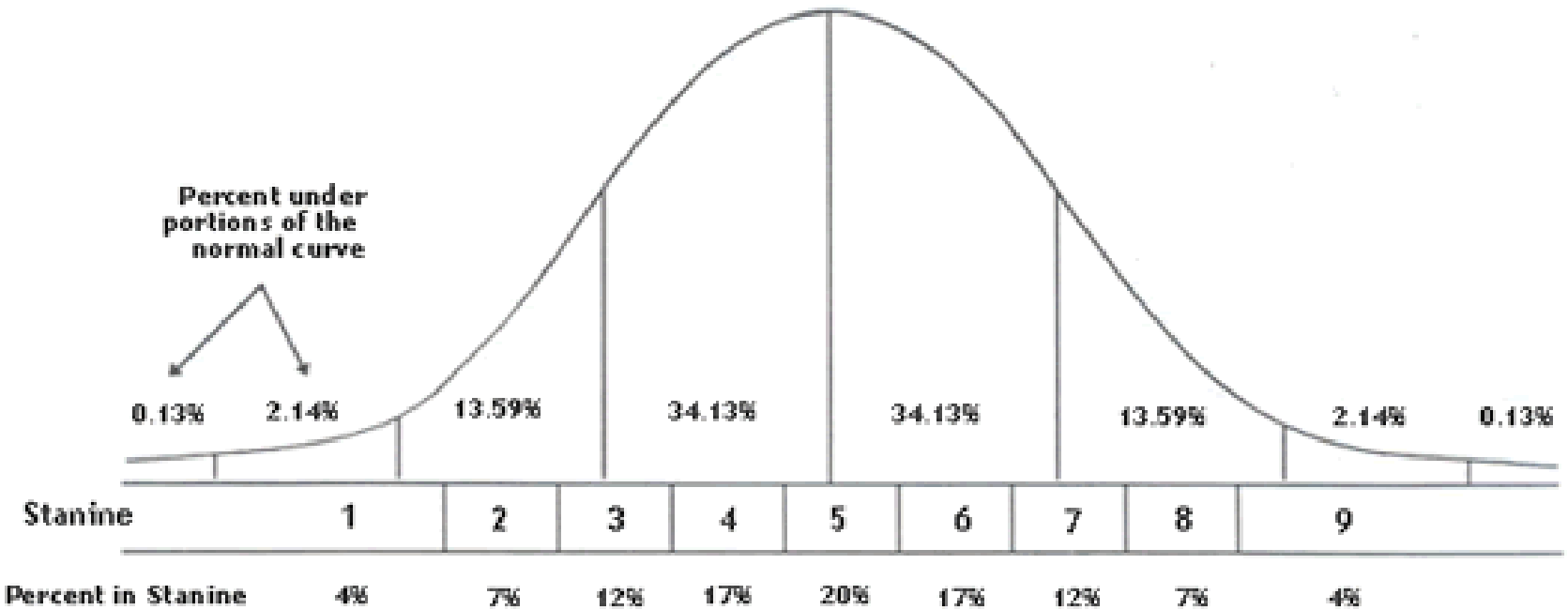
Issues with the Observation Survey

- Criticized for not being standardized
- Text level intervals are unequal
- Floor & ceiling effects on measures
- No bubbles to fill
- No profit; no marketing

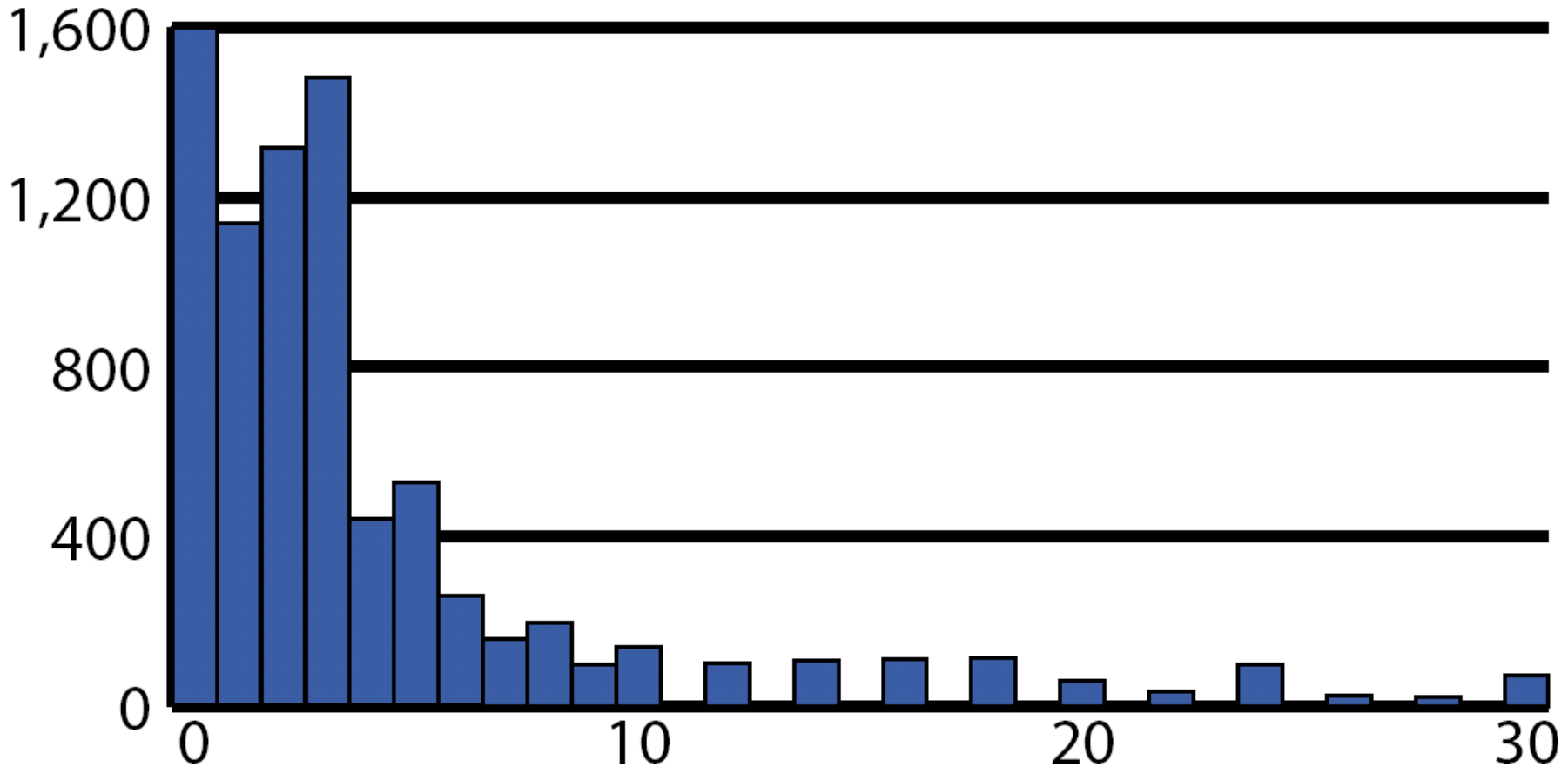
Normal Distribution

Figure 2

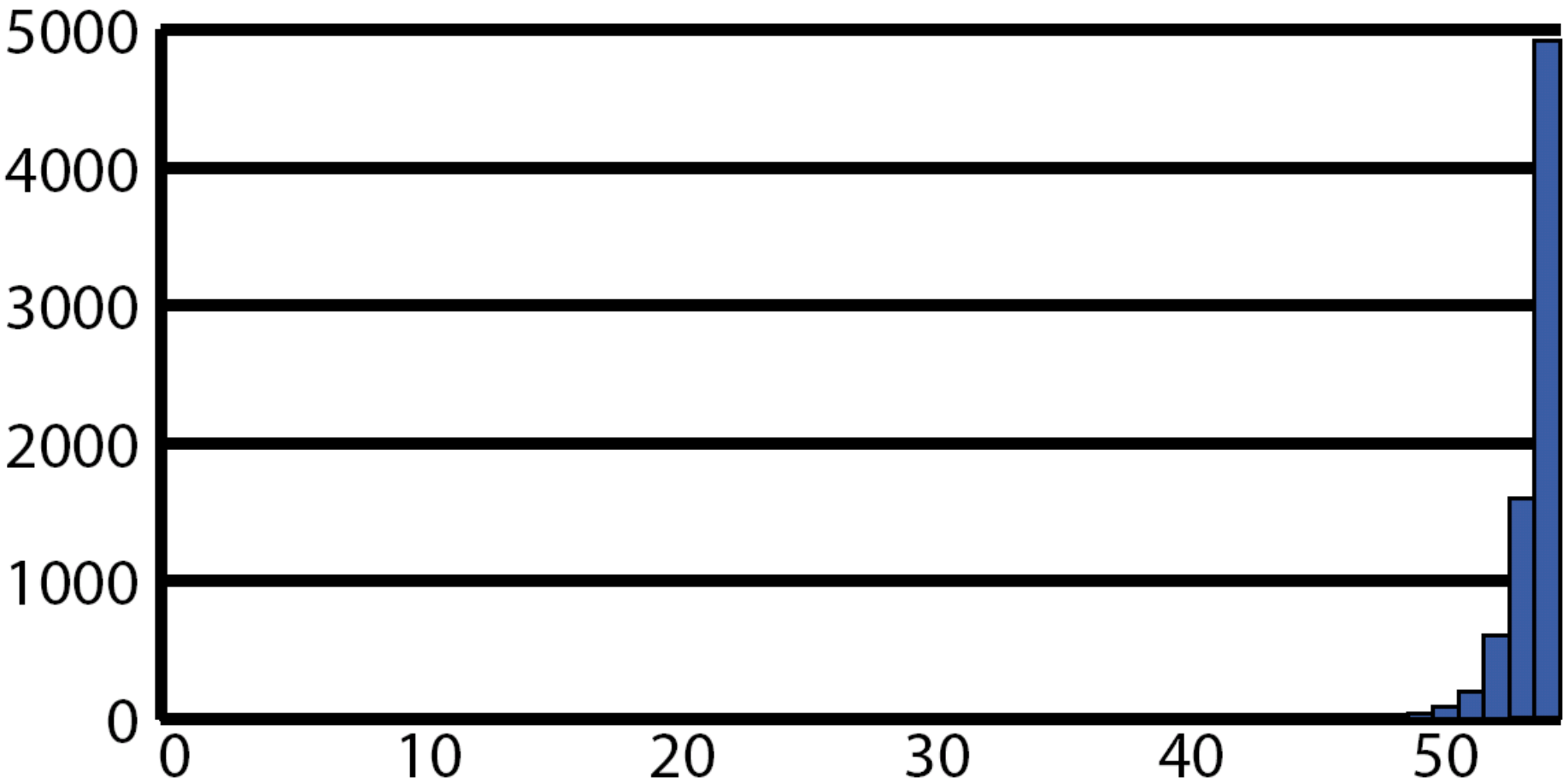
The dimensions contained in the evaluation are measured by the Stanine System. This system of measurement incorporates a line broken into nine standad sections. The nine Standardized sections can be compared to the Bell Curve of the general population, as shown in the illustration below. A score in the 1-3 stanine range equates to the bottom one-third of the population on the curve; a score in the 4-6 range equates to the mid-range of the population on the curve, and a score in the 7-9 range equates to the upper one-third of the population on the curve

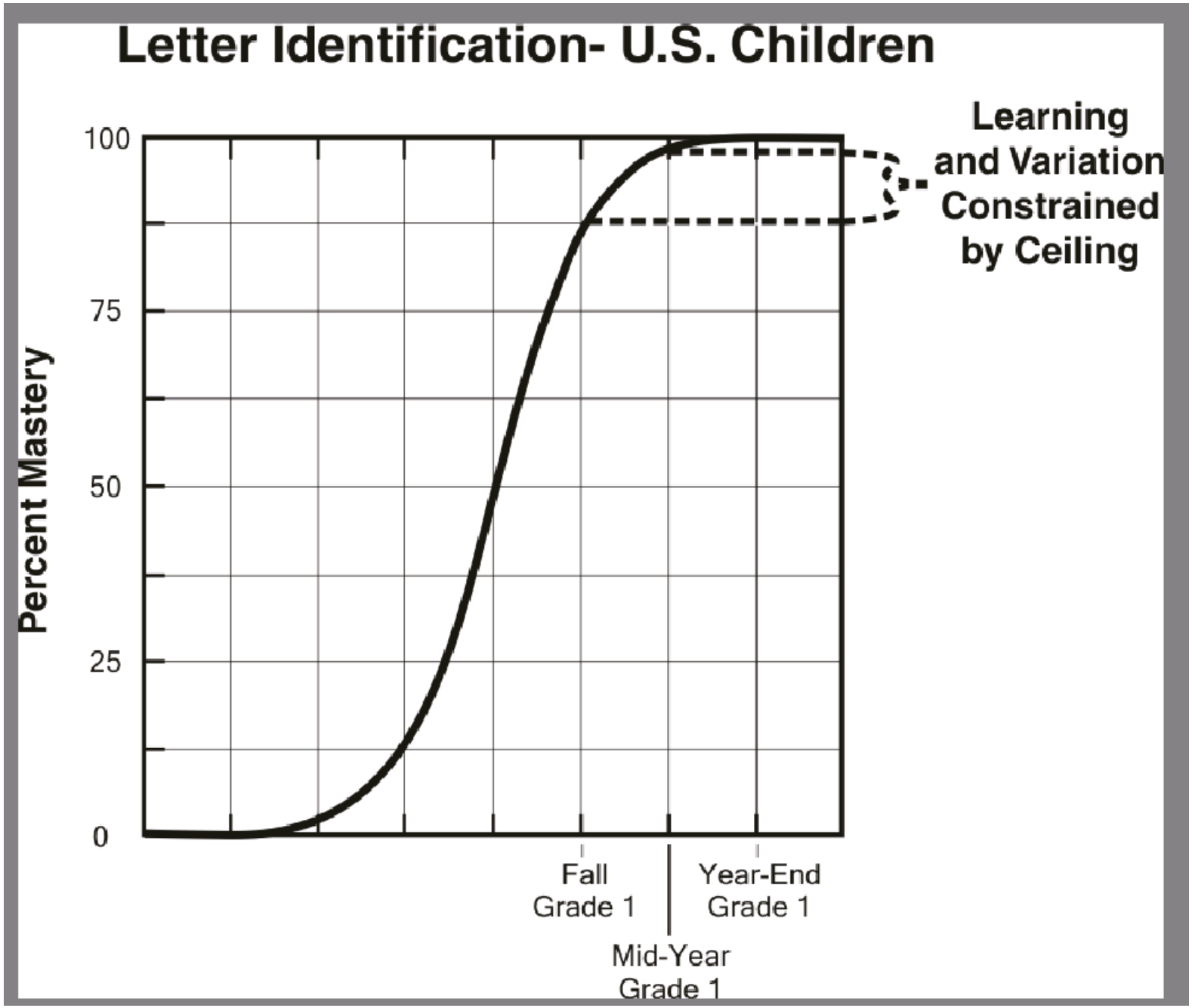


Text Reading Level-Fall



Letter ID-Spring




Letter Identification- U.S. Children
Learning and Variation Constrained by Ceiling
Percent Mastery
100
75
50
25
0
Fall Grade 1
Year-End Grade 1
Mid-Year Grade 1

Definitions

- Standardization means
  1. Administering a standard task
  2. Administering this task in a standard way (restrict conditions of observation)
- Standardization leads to reliability
- Reliability: Getting the same results under the same conditions
- Not the same as “objective” tests

Standard vs Valid & Reliable

- Validity: Does the assessment measure what it intends to?
- Letters and words are relatively easy
- Comprehension is very difficult
- An assessment can be standardized and reliable but not valid

Validity

- Validation involves examining evidence to support a proposed interpretation of a survey item
- Does the survey item measure what it is intended to measure?
  - Is this interpretation is justified?
  - Is this interpretation plausible?
  - Is this interpretation appropriate?

Purposes of Assessment

- Instruction
- Accountability
- Administrative decision making
- Research
- Different assessments may be needed for different purposes
- What is in the best interest of the student?

Purpose of Observation Survey

- Purpose of OS is to inform instruction, not to measure group change
- A criterion referenced assessment; same assessment given in fall and spring
- Statistical properties of OS are not an accident
- Less straightforward to analyze than normally distributed measures

Observation Survey-Iowa Test of Basic Skills Correlational Study

Francisco X. Gómez-Bellengé,
The Ohio State University
Emily Rodgers, The Ohio State University
Chuang Wang, University of North Carolina Wilmington
Melissa Schulz, University of Cincinnati
Paper Presented at 2005 Meetings of the
American Educational Research Association
Available for download at
http://www.ndec.us/Documentation.asp

Results

- Research Question 1: Does OS identify same students as low readers as ITBS?
- 90% of Reading Recovery students were identified as being among the lowest 20% readers by the ITBS
- Research Question 2: Do the subtests of the ITBS correlate with tasks of the OS in fall?
- RQ3: What are treatment effects

OS-ITBS Correlations-Fall

Items	LI	OWT	CAP	WV	HRSW	TRL
Vocabulary	.234**	.409**	.545**	.343**	.407**	.382**
Reading	.318**	.682**	.502**	.598**	.534**	.727**
Reading Total	.394**	.795**	.593**	.685**	.659**	.764**
Listening	.212**	.411**	.505**	.354**	.346**	.371**
Word Analysis	.476**	.603**	.660**	.533**	.623**	.534**
Language Usage	.223**	.463**	.533**	.439**	.430**	.428**
Language Total	.247**	.495**	.585**	.450**	.441**	.453**

** $p < .001$

Treatment Effect

	T	df	CI (99.6%)	P
Observation Survey				
LI	-15.174	244	-9.07, -6.98	.000*
OWT	-73.463	244	-16.21, -15.36	.000*
CAP	-39.645	244	-9.82, -8.89	.000*
WV	-45.734	244	-44.26, -40.60	.000*
HRSW	-56.684	244	-26.15, -24.39	.000*
TRL	-43.606	244	-16.54, -15.11	.000*
ITBS				
Reading	23.227	196	13.40, 15.89	.000*
Reading Total	25.626	196	13.64, 15.91	.000*
Language Usage	20.700	230	13.54, 16.39	.000*
Language Total	22.472	227	11.11, 13.24	.000*

Assessment and Identification of First Grade Students at Risk: Correlating the Dynamic Indicator of Basic Early Literacy Skills and An Observation Survey of Early Literacy Achievement

Francisco Gómez-Bellengé,
National Data Evaluation Center
Sharan A. Gibson,
San Diego State University,
Meiling Tang,
Ohio Department of Education,
Mary Anne Doyle,
University of Connecticut,
Patricia R. Kelly,
San Diego State University

Download from http://www.ndec.us/Documentation.asp

DIBELS & OS

Although DIBELS and OS are both intended to measure student progress in key aspects of early literacy acquisition and to assist with identification of at-risk students, the two assessments are based on significantly different theoretical constructs.

Purpose of Study

The purpose of this study was to determine correlation between subtests on the DIBELS and OS, correspondence in identification of beginning first-grade students at risk of reading failure, and instructional effects of the Reading Recovery early intervention (Clay, 1993) revealed by post-treatment testing on DIBELS and OS.

DIBELS

The Dynamic Indicators of Basic Early Literacy Skills (DIBELS) (Good & Kaminski) consists of a continuum of fluency-based indicators of early literacy skills that are foundational and intended to be predictive of reading success, with fluency defined as degree of accuracy and speed of responses.

Observation Survey

The Observation Survey reflects a multifaceted theory of beginning reading and writing with an emphasis on complex cognitive processing (Clay, 2002).

DIBELS Tasks

- Fall of Grade 1
  - Letter Naming Fluency
  - Phoneme Segmentation Fluency
  - Nonsense Word Fluency
- Spring of Grade 1
  - Phoneme Segmentation Fluency
  - Nonsense Word Fluency
  - Oral Reading Fluency

Observation Survey Tasks

- Letter Identification
- Ohio Word Test
- Concepts About Print
- Writing Vocabulary
- Hearing & Recording Sounds in Words
- Text Reading Level

Research Questions

1. Are students identified by Reading Recovery as “having the most difficulty learning to read” in fall also identified as “at risk” using the DIBELS?
2. Do specific measures from both instruments of OS and DIBELS correlate with each other in fall, at mid-year, and at year-end?

Research Question 1

Around 49% of Reading Recovery students were identified by DIBELS as “needs substantial intervention” based on five individual performance patterns. This finding suggests a substantial discrepancy between Reading Recovery and the DIBELS in terms of the identification of students at risk.

Research Question 2-Fall

Table 2

Correlations between the OS and the DIBELS in fall

Subtests	LI	OWT	CAP	WV	HRSW	TRL
LNF	**0.621****	0.541**	0.399**	0.537**	0.574**	0.422**
PSF	0.337**	0.366**	0.338**	0.375**	**0.508****	0.271**
NWF	0.455**	0.648**	0.405**	0.564**	**0.619****	0.559**
ORF	0.215**	*0.409***	0.114*	0.337**	0.267**	*0.382***
WUF	0.212**	0.266**	0.308**	0.278**	0.329**	0.217**
RTF	0.085	0.250**	0.240**	0.225**	0.232**	*0.373***

** $p<0.001$

* $p<0.05$

Research Question 2-Mid-Year

Table 3

Correlations between the OS and the DIBELS at mid-year

Subtests	LI	OWT	CAP	WV	HRSW	TRL
LNF	*0.206*	0.244*	0.2378	0.121	0.195	0.296**
PSF	0.227**	0.312**	0.247**	0.246**	**0.365****	0.303**
NWF	0.259**	0.468**	0.358**	0.348**	**0.342****	0.594**
ORF	0.266**	**0.637****	0.509**	0.506**	0.405**	**0.814****
WUF	0.141*	0.199**	0.161*	0.193**	0.201**	0.196**
RTF	0.220**	0.459**	0.431**	0.345**	0.408**	*0.568***

** $p<0.001$

* $p<0.05$

Research Question 2-Spring

Table 4

Correlations between the OS and the DIBELS at year-end

Subtests	LI	OWT	CAP	WV	HRSW	TRL
LNF	*0.083*	0.164**	0.144**	0.169**	0.123**	0.129**
PSF	0.085**	0.162**	0.129**	0.098**	**0.205****	0.151**
NWF	0.189**	0.351**	0.235**	0.266**	**0.293****	0.496**
ORF	0.234**	**0.483****	0.352**	0.429**	0.369**	**0.714****
WUF	0.061**	0.153**	0.148**	0.108**	0.153**	0.173**
RTF	0.120**	0.334**	0.301**	0.304**	0.281**	*0.439***

** $p<0.001$

* $p<0.05$

Research Question 3

Do subtests of DIBELS and corresponding observation tasks of OS show similar treatment effects for each student group?

Research Question 3

The examination of fall to spring differences on the PSF and NWF scores attained by RR students revealed significant differences, with Wilks' Lambda value of 0.183. The RR students attained significantly higher scores on the PSF and NWF tests in the spring. This suggests that the Reading Recovery intervention made a difference for the RR students as revealed by DIBELS' spring scores

Discussion/1

This study's findings confirm low to very high correlations between subtests on the Dynamic Indicators of Basic Early Literacy Skills and An Observation Survey, as well as significant treatment effects for RR students on both DIBELS subtests and OS. A problematic finding, however, is the difference in identification of at-risk students between DIBELS and OS.

Discussion/2

Differences in the identification of at-risk learners between the DIBELS and the OS are related to the alternative theories of reading acquisition inherent in the two measures. DIBELS focuses on discrete skills considered predictive of future success (statistically determined). The OS focuses on both specific skills and more complex processing strategies ascribed to mature reading and writing.

Discussion/3

More specifically, while the DIBELS appears to assume that the acquisition of discrete skills (e.g., the alphabetic principle) causes later success; the OS is constructed to allow assessment of the complexity of literacy behaviors (reading and writing behaviors) considered paramount from the earliest acquisition phase and necessary for ongoing, successful performance and learning.

Issues with DIBELS/1

- Misperception that it was required by Federal requirement
- Misperception that DIBELS is an intervention
- Conflict of interest problems
- See Inspector General Reports
- Problem with classification criteria

Classification criteria

- Several articles point out problems with the criteria used in DIBELS system for identifying students as at risk or not
- See references in paper
- DIBELS may under-identify students who do OK on specific tasks but are not putting it all together & processing text

Lessons from Voyager

- Latest review from What Works Clearinghouse says that Voyager intervention produces a large ***negative*** effect on comprehension
- Voyager developed by many of the same people that developed DIBELS
- Emphasis on phonics, speed & on mechanics may be at the detriment of comprehension (processing text)

Implications for Policy

- Results from NDEC reports can be relied on; gains shown on OS would also be shown on other instruments
- There is no need for routine external validation of OS results
- Since RR is effective and the OS is reliable and valid, local decision makers need to focus on local results with a focus towards improving implementation.