COMPARING THE EFFECTIVENESS OF UNIVERSAL SCREENING MEASURES OF READING

David A. Klingbeil & Anne F. Zaslofsky
PRESS: Path to Reading Excellence in School Sites

PARTNERSHIP AMONG:

- Minneapolis Public Schools
- UMN/Minnesota Center for Reading Research
- Minnesota Reading Corps
- Target
- Harvest Preparatory
- Best Academy & Best Academy East
Overview

- Purposes of Universal Screening
- Steps to Selecting Screening Measures
- Desiderata for Screening Measures
- Our study:
  - Methods
  - Results
  - Implications/Limitations/Future Research
- Discussion
Introduction

Background for our study
Quick Poll: Are these students at-risk?

Nora
- F&P = 3 levels below
- CBM-ORF = 8th percentile
- MAP = 7th percentile

Jayda
- F&P = at grade level
- CBM-ORF = 20th percentile
- MAP = 24th percentile

Mac
- F&P = at grade level
- CBM-ORF = 71st percentile
- MAP = 20th percentile

Get out your phones!
Text your response to 22333
- Nora: 223416
- Jayda: 223418
- Mac: 223458

Poll Results
Universal Screening

“Universal screening is at the center of the RTI process” – Shapiro & Gebhardt, 2012

- Administered to all students
- Aligned with curriculum
- General outcome measure (GOM)
- Delivered several times a year

Deno, 2003; Ikeda et al., 2007
Universal Screening

- Questions that can be answered include:
  - Is there a school- or class-wide problem?
  - Which students are at-risk for future problems?
  - Are our current intervention efforts working?

- Apply to academic and behavioral domains

Glover & Albers, 2007; Fuchs, Fuchs, & Compton, 2012
Critical Characteristics

- Screening is aligned with the curriculum
  - Especially important in math
  - Appropriate for what your school needs.

- Otherwise unreliable

- Private School Example

Ardoin et al., 2004
Critical Characteristics

- Technically Adequate
  - Local or National Norms?

- Reliability

- Validity
  - Concurrent & Predictive

Hosp, Hosp, & Dole, 2011; Ikeda et al., 2008
Critical Characteristics

Predictive Validity

- Most important characteristic of screening measure

- Four possible outcomes

Sensitivity & Specificity

<table>
<thead>
<tr>
<th>TP</th>
<th>TN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correctly identified as at-risk</td>
<td>Correctly identified as not at-risk</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FP</th>
<th>FN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorrectly identified as at-risk</td>
<td>Incorrectly identified as not at-risk</td>
</tr>
</tbody>
</table>
Critical Characteristics

- Should be easy to administer, score objectively
- Minimal time away from instruction
- Cost-benefit ratio
Comparing Two Screening Measures

Methods & Results
Comparing Two Screening Measures

Research Questions

1. Which measures predict the most variance in Measures of Academic Progress scores?
2. Which combination of measures results in highest predictive validity?
3. What is the effect of decision rules on SE? SP?
Comparing Two Screening Measures

Method:

- Participants –
  - 609 2\textsuperscript{nd} and 3\textsuperscript{rd} grade students
  - approximately 80\% ethnic minority
  - \approx 75\% FRL

- 4 schools in large urban district
Measures

CBM-R

AIMSweb curriculum
- *Reliability of .90-.96
- Moderate correlations with standardized tests and state assessments
- Varied results for predictive validity

Benchmark Assessment System

Heinemann Publishing
- Test-Retest - .93
- Convergent validity - .44-.94
- No studies of predictive validity
Measures

CBM-R
AIMSweb curriculum
- $4.00 per student
- Printing costs
- Three, 1-min reads
- WRC/E

F&P Benchmark Assessment System
- K-8 BAS costs $760
- Online system costs $30 per classroom
- 15-30 min per student
- Level and Comp
Outcome Measure

Measures of Academic Progress (NW Eval. Assoc.)

- Computer Adaptive -
- Aligned with several state tests
- Test-Retest Reliability .90-.95
- Concurrent Validity - .80
- 40-60 minutes per student
- Approximately $14.00 per student
Procedures

- Students completed each measure in Sept., Jan., and Apr.
- Graduate students administered the ORF measures, classroom teachers administered the BAS.
- Students completed MAP assessments per district standards.
Analytic Plan

Question 1
- Hierarchical OLS regression models to predict Spring MAP scores.

Question 2
- Receiver Operating Curves
- Logistic Regression for multiple screeners
Regression Results

- Each predictor was significantly correlated with each other and the outcome.

Step 1:
- Spring MAP ~ Fall ORF Z scores
- Spring MAP ~ Fall BAS

Step 2:
- Spring MAP ~ Fall ORF
  - ~ Fall ORF + Fall BAS
  - ~ Fall ORF + Fall BAS + Fall MAP
Additional Models

What are we adding? Is Fall MAP enough?

Step 2:

Spring MAP $\sim$ Fall MAP

$\sim$ Fall MAP + Fall ORF

$\sim$ Fall MAP + Fall ORF + Fall BAS
Regression Results

Step 1:
- Fall ORF: $r^2 = 0.44; p < 0.001$
- Fall BAS: $r^2 = 0.41; p < 0.001$

Step 2:
- Fall ORF: $r^2 = 0.46; p < 0.001$
- Fall ORF + BAS Both $p < 0.001$
  - $r^2 = 0.56$
- Full
  - All 3 $p < 0.001$
  - $r^2 = 0.74$
Regression Results

Final model:
Spring MAP ~ Fall ORF + Fall MAP

Both predictors significant ($p < .001$), $r^2 = .73$

BAS requires teachers to spend an extra 10-15 min per student but predicts similar variance to Fall MAP + ORF.
Receiver Operating Curves

- Judges diagnostic accuracy
- Graphs SE and SP
- Area Under the Curve = effect size
Diagnostic Efficiency

Fall ORF Z
- AUC = .82
- SE = .67
- SP = .81

Fall BAS
- AUC = .74
- SE = .72
- SP = .69

Missed 77 students!!!
What do we do?
Previous researchers combined measures.
Diagnostic Efficiency

How?

We could use visual criteria.
Diagnostic Efficiency

ORF + BAS
- SE = .71
- SP = .79

Decreased False Negatives by 2% from ORF.

Decreased False Positives by 5.5% from BAS.

Recall:

ORF
- SE = .67
- SP = .81

BAS
- SE = .67
- SP = .81
Diagnostic Efficiency

Full Model
SE = .76
SP = .83

Recall:

ORF
SE = .67
SP = .81

BAS
SE = .67
SP = .81

ORF + BAS
SE = .67
SP = .81

Visual helps but did not improve to recommended SE level.
Diagnostic Efficiency

Visual criteria did not meet recommended SE levels.

We could use logistic regression.
Diagnostic Efficiency

Logistic regression requires more steps:

1. Recode Spring MAP into at-risk/not at-risk variable
2. Create Spring MAP \sim\ Predictors
3. Save predicted probabilities
4. Enter predicted probabilities into ROC analysis
5. Use coordinates to determine cut score where SE = .90

Catts et al., 2001; Johnson et al., 2010
Okay now what?!

Cannot wait until data is collected all year to determine risk.

Catts et al. (2001) highlights how to create a risk calculator using excel.

Risk = 1/(1/ β₁ * ORF + β₂ * BAS, + β₃ * MAP)
Diagnostic Efficiency

**BAS + ORF**

SP = .54 (decreased from .79)
False Positives: 11.5% ➔ 25.5%

**Full**

SP = .69 (decreased from .83)
False Positives: 8.7% ➔ 16.9%

SE held at .90 ≈ 23 false negatives
Results

- ORF and BAS alone predicted about 40% of variance in Spring MAP scores.

- Combination of the two measures predicted 10% more variance.

- ORF + MAP has similar results as full model. Fewer resources.
Results

More than 1 measure helps, but visual inspection does not meet recommended SE criteria (missing too many at risk).

Logistic regression allows us to hold SE constant but increases number of false positives.

Full: interventions for 57% of students ($n = 258$), 75 didn’t need it.
Discussion

Implications for Practice & Limitations
Quick Poll: Are these students at-risk?

- Nora
  - F&P = 3 levels below
  - CBM-ORF = 8\textsuperscript{th} percentile
  - MAP = 7\textsuperscript{th} percentile

- Jayda
  - F&P = at grade level
  - CBM-ORF = 20\textsuperscript{th} percentile
  - MAP = 24\textsuperscript{th} percentile

- Mac
  - F&P = at grade level
  - CBM-ORF = 71\textsuperscript{st} percentile
  - MAP = 20\textsuperscript{th} percentile

- Get out your phones!
- Text your response to 22333
  - Nora: 223416
  - Jayda: 223418
  - Mac: 223458

Poll Results
Limitations

- Attrition and missing data
  - Urban district some is expected

- Generalizability

- Nested data
Implications

- Combination of measures appears to improve from ORF or BAS alone.
  - BAS also provides some instructional information but we did not consider those effects.

- Consider the resources spent screening
  - CBM-R: 3-5 min, $0 to $4 per student
  - BAS: 15 min, ≈$2 per student
  - Constructs measured
  - Subjective/Objective
Conclusions

- School psychologists should help guide screening decisions

- These techniques are straightforward ways to examine combinations of screening measures

- When technical adequacy is similar, look at aesthetic characteristics
Questions

Thank You!!!

Contact info:

Dave Klingbeil – kling124@umn.edu
Anne Zaslofsky – azaslofs@umn.edu
Matt Burns – burns258@umn.edu


