Relations between CBM (Oral Reading and Maze) and Reading Comprehension on State Achievement Tests: A Meta-Analysis

Shin, Jaehyun PhD
Department of Special (Inclusive) Education
Gyeongin National University of Education
Overview

• Background
• Problem Statement
• Purpose and Research Questions
• Method
• Results
• Discussion
• Q&A
Background

• Emphasis on accountability for success in reading
  – Need screening tools to identify students at-risk for reading failure on state achievement tests

• Many school districts implement Curriculum-Based Measurement (CBM, Deno, 1985) as screening tools
  – Valid indicator of ‘overall reading proficiency’ (Wayman et al., 2007; Wanzek et al., 2009)
  – Oral reading and Maze received the most attention
My Dress-Up Box

I like to pretend, do you? We have a dress-up box at our house. I use it when I pretend. There are clothes and hats and shoes in it. There are some old Halloween costumes.

The dress-up box is like magic. It helps me be anything I want. I can pretend I am in the army. All I do is put on some army clothes and my dad’s old hat. I pretend to be a fireman when I wear a yellow coat and black boots.

Sometimes my friend Eric comes over. We both pretend we are cowboys or robots. Sometimes we explore the back yard. We pretend it is the first time anyone has seen it. Sometimes we find new treasures.

My sister Lisa likes to pretend, too. She puts on a long purple dress and a gold crown. Then she plays like she’s queen of the world. She was a pretty bossy queen.

Once she put on some gloves and a hat. She carried one of my grandma’s purses. She rang our doorbell. When my mother opened the door, Lisa said she was our new neighbor. So my mother pretended, too. She asked if Lisa wanted to come in for tea.

Playing dress-up is fun.

Error Pattern: Total: __________
SUMMER CAMP

Stuart had nice parents. They did not embarrass him in [glad/ front/ yellow] of his friends. His father did [not/ ant/ soft] yell at him during his baseball [center/ games/ lines], and his mother never kissed him [in/ tot/ put] front of his friends. He generally [liked/ flow/ jeep] his parents, except for the fact [shoe/ went/ that] they were sending him to summer [bus/ dump/ camp] this year.

Stuart did not want [to/ with/ cow] go to summer camp. The thought [and/ be/ off] it made him picture himself hot [coat/ rest/ and] thirsty, hiking up a dusty trail. [Bit/ He/ Go] knew that summer camp food had [off/ to/ my] be bad news, too. Besides, summer [camp/ free/ dog] was for people with nothing else [fad/ to/ sew] do. He had plenty of things planned [for/ much/ very] his summer at home.

"Summer camp [will/ yes/ belt] be good for you," said Mother. "[Feel/ And/ Lot] I don’t want to hear another [catch/ phone/ word] about it!" Stuart moped around the [beat/ opens/ house] until it was time to go. Mother [had/ with/ boy] packed his trunk full of clothes, [and/ sort/ time] she and Dad took Stuart to [real/ glob/ the] bus station. Stuart tried hard not [to/ sun/ we] cry when he hugged them goodbye. [Yet/ He/ Sat] ran onto the bus and buried [beam/ his/ neat] head in his hands. After a [while/ tall/ hate], he looked out the window.

Camp Mac [was/ snow/ rent] in the Cheaha Mountains of Alabama. Stuart [twist/ never/ girls] knew there were any mountains in Alabama. [The/ Now/ Man] bus climbed and climbed and soon stopped. [A/ By/ In] man helped Stuart carry his trunk [lot/ to/ yes] the camp gate. There he was [met/ when/ ripe] by a teenage boy with a [while/ risen/ suntan] and a whistle around his neck. "[In/ Hi/ Or] there! You must be Stuart Sikes. I’m Tom, [here/ into/ your] cabin counselor. I’ll help you get settled [in/ go/ dot], then we will go eat lunch [serve/ with/ goal] the others."

Stuart was quiet and followed Tom. [There/ Either/ Ferry] were so many children at the [will/ camp/ she], and they all were having fun. [Box/ They/ Lane] looked at Stuart and someone said [hello/ pretty/ last]. Stuart was already homesick and his stomach [ice/ book/ hurt].

After lunch they all went down [by/ top/ out] the lake to go for canoe [faith/ still/ ride]. Three boys asked Stuart to join [them/ sent/ jeans] and Stuart did with a smile. [Hurt/ Trip/ Maybe] they would be his friends.

By [that/ know/ lope] night Stuart had forgotten about Mother [so/ step/ and] Dad. He was having so much [nose/ fun/ body] at summer camp that he did [bet/ not/ mad] want to be any place else.
Wayman et al. (2007) review on CBM reading

<table>
<thead>
<tr>
<th></th>
<th>Oral reading</th>
<th>Maze</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face validity</td>
<td>Relatively low</td>
<td>Relatively high</td>
</tr>
<tr>
<td>What they measure</td>
<td>More than just a measure of fluent decoding</td>
<td>Multiple aspects of reading</td>
</tr>
<tr>
<td>Criterion validity</td>
<td>Moderate to strong for primary grades (1-3), but</td>
<td>Correlations remain fairly stable across</td>
</tr>
<tr>
<td></td>
<td>tends to decline in intermediate grades</td>
<td>elementary grades</td>
</tr>
</tbody>
</table>
Problem Statement

• Much of validity work targeted students in Grades 1-3 who often struggle with decoding/fluency
  – Less work targeted older students who may have reading comprehension (RC) difficulties

• Despite maze appears to be a better indicator of RC than oral reading does especially for older kids, empirical evidence is insufficient (Wayman et al., 2007)
  – More evidence is needed on the validity of oral reading and maze as indicators of RC across grades
Purpose and Research Questions

• Integrative conclusion needed for
  – Relative validity of oral reading/maze across grades
  – Potential moderator effects (e.g., grade level, characteristics of CBM/state test)

• Use of meta-analytic approach
  – Converging evidence beyond previous qualitative, quantitative reviews (Wayman et al., 2007; Reschly et al., 2009; Yeo, 2008; 2010)
Purpose and Research Questions (cont.)

- Research Questions
  1) What are the estimated average correlations between CBM (oral reading, maze) and RC on state tests, and do they vary by (a) type of task and (b) grade level (primary, intermediate, secondary)?
     - H1: correlation for oral reading will be larger for primary grades
     - H2: correlation for maze will be larger for intermediate, secondary grades
  2) To what extent are the relations between CBM (oral reading and maze) and RC on state tests influenced by potential moderating factors?
METHOD
Literature Search

• Search procedure
  – Database (ERIC, PsycINFO, Google Scholar) plus CBM websites (e.g., AIMSweb, easyCBM)
  – Ancestral searches & hand search

• Inclusion criteria
  1) Examine relation between CBM and RC on state tests
  2) CBM administered before or concurrently with state test
  3) Sufficient quantitative information
  4) Participants had to be students in Grades 1-12
  5) Only correlational or group experimental designs
## Coding System

<table>
<thead>
<tr>
<th>Category</th>
<th>Definition of descriptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants</td>
<td>Grade range (Grades 1-3; Grades 4-6; and secondary)</td>
</tr>
<tr>
<td></td>
<td>Student demographic (female, ELL, FRL, Sped, White %)</td>
</tr>
<tr>
<td>CBM</td>
<td>Type of CBM (oral reading or maze)</td>
</tr>
<tr>
<td></td>
<td>Number of passaged used</td>
</tr>
<tr>
<td></td>
<td>Development type (commercial vs researcher-developed)</td>
</tr>
<tr>
<td></td>
<td>Administrator type (researcher vs school personnel)</td>
</tr>
<tr>
<td>State test</td>
<td>Development type (commercial vs state-developed)</td>
</tr>
<tr>
<td></td>
<td>Response format (multiple choice vs mixed format)</td>
</tr>
<tr>
<td></td>
<td>Time interval (between CBM and state test; in month)</td>
</tr>
<tr>
<td>Study</td>
<td>Type of publication (journal, dissertation, tech report)</td>
</tr>
</tbody>
</table>
Search and Coding Reliability

• Search IRA
  ‒ Approx. 20% of the studies identified were randomly selected and screened by a doc student in EPsy
  ‒ IRA = 84% (all disagreements were resolved)

• Coding IRA
  ‒ I served as the primary coder
  ‒ Approx. 20% of all studies were coded by a doc student in C&I as a 2\textsuperscript{nd} coder who was trained on definitions of variables
  ‒ IRA = 94.1% (all disagreements were resolved)
Meta-Analytic Procedure

- Random effects, two-level model
  - Level-1: mean ES & sampling error
  - Level-2: variability in ES among studies

- Procedure
  1) Unconditional model
  2) Conditional model
     1) Preliminary moderator analysis (one-moderator-at-a-time)
     2) Meta-regression
Final Studies/Correlations Included

• Of 61 studies with 237 correlations, 105 correlations excluded due to dependency issue (Harwell & Maeda, 2008)
  – e.g., same kids in a study administered CBM multiple times
  – Only 1 correlation was included (most concurrent, closest relations) in the final analyses

• Addressing outlier
  – 1 study seemed to be a potential outlier
  – Overall correlation and $I^2$ were similar with and without it
RESULTS
Characteristics of Studies

• Study: Journal article (62%), dissertation (25%), tech report (13%)

• Participants
  – About 80% of studies on elementary grades
  – Mean proportion: ELL (13%), Sped (14%), FRL (49.5%)

• CBM
  – 80% of studies focused on oral reading
  – 84% of studies used commercial CBM (e.g., DIBELS, easyCBM)

• State Test of RC
  – Multiple choice & mixed-format were equally used
  – 85% of studies administered state tests within 6 months after CBM
Unconditional Analysis

• **Average correlations**
  – All CBM combined (n=132): $r = .63$ (95% CI = .62-.64)
  – Oral reading (n=103): $r = .63$ (95% CI = .60-.66)
  – Maze (n=29): $r = .60$ (95% CI = .57-.62)

• **Heterogeneity**
  – Sig. Q-values: true ES varies from study to study
  – $I^2$ (between-study heterogeneity): 98.91 (all), 98.74 (oral reading), 99.28 (maze), indicating that about 99% of the observed variance reflects ‘*real differences*’ in study effects
Moderator Analysis: one-at-a-time

As a preliminary analysis, a series of traditional moderator analyses were done for each categorical moderator:

- Type of publication
- Grade range
- Student demographics
- Type of CBM
- Development type of CBM
- Number of CBM passages
- Type of CBM administrator
- Development type of state test
- Response format of state test
- Time interval between CBM and state test
Moderator Analysis: one-at-a-time (cont.)

- Significant moderators were as follows:

<table>
<thead>
<tr>
<th>Moderator</th>
<th>$p$</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade range (for both CBMs)</td>
<td>.016</td>
<td>Intermediate, primary, secondary</td>
</tr>
<tr>
<td>Grade range (for maze)</td>
<td>.003</td>
<td></td>
</tr>
<tr>
<td>Type of CBM (all grades combined)</td>
<td>.008</td>
<td>Oral reading, maze</td>
</tr>
<tr>
<td>Type of CBM (for intermediate &amp; secondary grades combined)</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td>Sped % (categorized)</td>
<td>.000</td>
<td>low, medium, high proportion</td>
</tr>
<tr>
<td>FRL % (categorized)</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>Time interval (categorized)</td>
<td>.001</td>
<td>1 month, 1-6 months, over 6 months</td>
</tr>
</tbody>
</table>

*Note. alpha = .05*
Moderator Analysis: meta-regression

• To examine (a) the effect of each moderator when other moderators held constant (b) the amount of variance accounted for by the set of moderators

• Two meta-regression models were analyzed due to missing data in student demographics
  – Model 1: all moderators except for student demographics
  – Model 2: added demographics to Model 1 \((n=35)\)
### Moderator Analysis: meta-regression (cont.)

<table>
<thead>
<tr>
<th>Sig. moderator at $\alpha=.05$</th>
<th>Model 1</th>
<th>Model 2 ($n=35$, for demographic)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of publication</td>
<td>Journal, tech report &gt; dissertation</td>
<td>Female %</td>
</tr>
<tr>
<td>Grade range</td>
<td>Primary, intermediate &gt; secondary</td>
<td>Sped %</td>
</tr>
<tr>
<td>Type of CBM</td>
<td>Oral reading &gt; maze</td>
<td>FRL %</td>
</tr>
<tr>
<td>Development type of state test</td>
<td>State-developed &gt; commercial</td>
<td></td>
</tr>
<tr>
<td>Time interval between CBM and state test</td>
<td>Shorter interval, higher correlation</td>
<td>*Other moderators included in Model 1 were significant, except for type of CBM</td>
</tr>
</tbody>
</table>

| $R^2$                         | .28    | .59    |
Publication Bias

Funnel Plot of Precision by Fisher’s Z

Precision (1/Std Err)
Fisher’s Z
DISCUSSION
Validity of Oral reading vs. Maze

- Average correlation
  - $r = .63$ (oral reading), $r = .60$ (maze) \((large, Cohen, 1992)\)
  - Valid indicators of RC on state tests across grades

- Sig. effect of type of CBM
  - Oral reading appears to be slightly more predictive of RC than maze does, especially for older kids
Validity of CBM by grade range

- Average correlations sig. differed by grade range
  - For both CBM: sig. lower correlations for secondary grades
  - For maze: correlation decreased considerably from intermediate to secondary grades

- Validity of CBM (especially for maze) may decrease at the secondary level
  - Contradict to previous findings that maze may be a better indicator of RC for older students
Validity of CBM by grade range (cont.)

- Present meta-analysis suggest oral reading might be slightly better indicator of RC than maze across grades (even for intermediate & secondary grades)
  - Automatic information processing theory (LaBerge & Samuels, 1974): *automatic decoding frees up cognitive resource for RC*
  - Oral reading fluency may reflect efficient text processing (Denton et al., 2011)
  - Perhaps maze only measure sentence level comprehension (Carlson et al., 2014; January & Ardoin, 2012)
Other Moderating Effects

• Not-significant, but noteworthy moderating effects
  – Development type of CBM (commercial, researcher-developed)
  – Number of CBM (single or mean, median scores)
  – CBM administrator type (researcher, school personnel)
  – Response format of state tests (multiple-choice, mixed)

• CBM as valid indicator of RC across setting or condition
Limitations and Future Research

• Limitations
  – Grade level had to be categorized as grade range due to very small sample size for some grades
  – Effects of demographic info should be interpreted with caution

• Future directions
  – Other potential moderators
  – Further research on maze
  – Use of CBM for a specific group of students
  – More research for secondary level
  – Research on ‘multiple-gating’ approach
Implications for Practice

• Converging evidence about CBM as valid indicators to assess RC across grades, with a bit differential validity by CBM type and grade level
  – Better indicator for elementary students than secondary
  – Correlations were “large” for both CBM: Practitioners may use oral reading and maze interchangeably

• Oral reading as valid indicators of RC despite low face validity

• Flexibility and consistency of CBM across various settings (e.g., materials, administrators)
Email: edusjh01@ginue.ac.kr

THANK YOU !!
본 연구의 목적은 읽기 학습장애 위험군 선별검사로 널리 활용되고 있는 두 가지 교육과정중심측정(CBM) 검사-구두읽기검사와 선택형읽기검사와 주(state) 성취도검사로 측정한 읽기이해와의 관계를 메타분석을 통해 살펴보는 것이었다. 이를 위하여 1~10학년을 대상으로 한 총 61편의 연구(132개 상관관계)가 최종적으로 분석에 포함되었다. 또한 평균 효과크기(상관계계)의 추정 및 잠재적인 조절변인 효과(연구특성, 샘플특성, CBM특성, 성취도검사특성 등) 분석을 위해 무선효과 메타분석을 적용하였다. 분석결과 CBM유형에 따른 상관관계 차이에 있어서는, 모든 학년을 합친 경우 구두읽기검사의 평균 상관관계가 선택형읽기검사의 평균 상관관계보다 유의하게 높았다. 학년수준을 나누어서 분석한 결과, 초등4학년 이후에서만 두 검사간의 평균 상관관계에서 유의한 차이를 보였으며 구두읽기검사의 상관관계가 조금 더 높게 나타났다. 다음으로 학년수준에 따른 CBM과 성취도검사로 측정한 읽기이해 간의 상관관계의 차이에 있어서는 두 검사가 유사한 패턴을 보였다. 구체적으로, 초등학교에서는 구두읽기검사와 선택형읽기검사 모두에서 학년수준에 따른 평균 상관관계가 유의하게 다르게 나타나지 않은 반면 중학교 이후에서는 초등학교에 비해 상관관계가 하락하는 경향을 보였다. 이같은 CBM 유형과 학년수준별 차이 외에도, 연구물 유형, 성취도검사 개발유형(상업용 또는 주 자체개발), CBM과 성취도검사 시행간 시간간격 등이 상관관계의 유의한 차이에 영향을 주는 변인들이었다. 이러한 연구결과를 바탕으로 읽기 학습장애 위험군 선별과 관련한 연구 및 실제에 주는 함의를 논의하였다.

주제어: 교육과정중심측정(CBM), 구두읽기검사, 선택형읽기검사, 읽기이해, 타당도, 메타분석