


# Content coverage and other measures of L2 speaking and listening proficiency in integrated tasks

William Bonk, Saerhim Oh, Masa Suzuki

Pearson

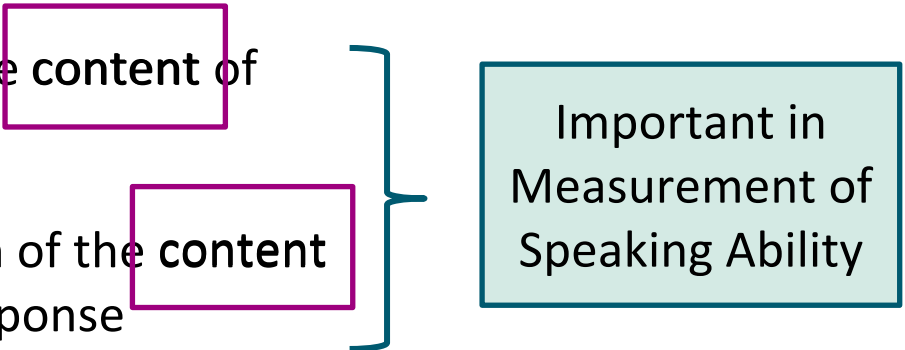
October 13, 2018

# Background of Study

- Many proficiency tests for young ELLs include integrated speaking tasks involving **listening** to a text and then **retelling** or **summarizing** it
  - How well ELLs understand the content of the listening material
  - How well the comprehension of the content is reflected in the spoken response
- 
- Important in Measurement of Speaking Ability
- Despite its importance, little research on understanding how content of source materials is reflected in young ELLs' speech (Hsieh & Wang, 2017)

# Background of Study

How do ELLs of different grade bands and proficiency levels represent the content of listening passages in their spoken responses?

- How well ELLs understand the **content** of the listening material
  - How well the comprehension of the **content** is reflected in the spoken response
- 
- Important in Measurement of Speaking Ability
- Despite its importance, little research on understanding how content of source materials is reflected in young ELLs' speech (Hsieh & Wang, 2017)

# Background of Study

- Many proficiency tests for young ELLs include integrated speaking tasks involving listening to a text and then retelling or summarizing it

What is the relationship between L2 oral proficiency and the ability to perform Listen-Retell tasks?

is reflected in the spoken response

- Despite its importance, little research on understanding how content of source materials is reflected in young ELLs' speech (Hsieh & Wang, 2017)



# Research Questions

How do ELs of different grade bands and proficiency levels represent the content of listening passages in their spoken responses?

1. Do ELLs at varying levels of L2 oral proficiency represent different **amounts of content** in their listen-retell tasks?



# Research Questions

How do ELs of different grade bands and proficiency levels represent the content of listening passages in their spoken responses?

2. Do ELLs of varying levels of L2 oral proficiency demonstrate difference in levels of performance on **content** and **language use**?



# Research Questions

What is the relationship between L2 oral proficiency and the ability to perform Listen-Retell tasks?

3-a. What do the measurement models tell us about the variables and their relationships?

3-b. What does the structural portion of the model tell us about the relationship between L2 Oral Proficiency and L2 Listen & Retell?

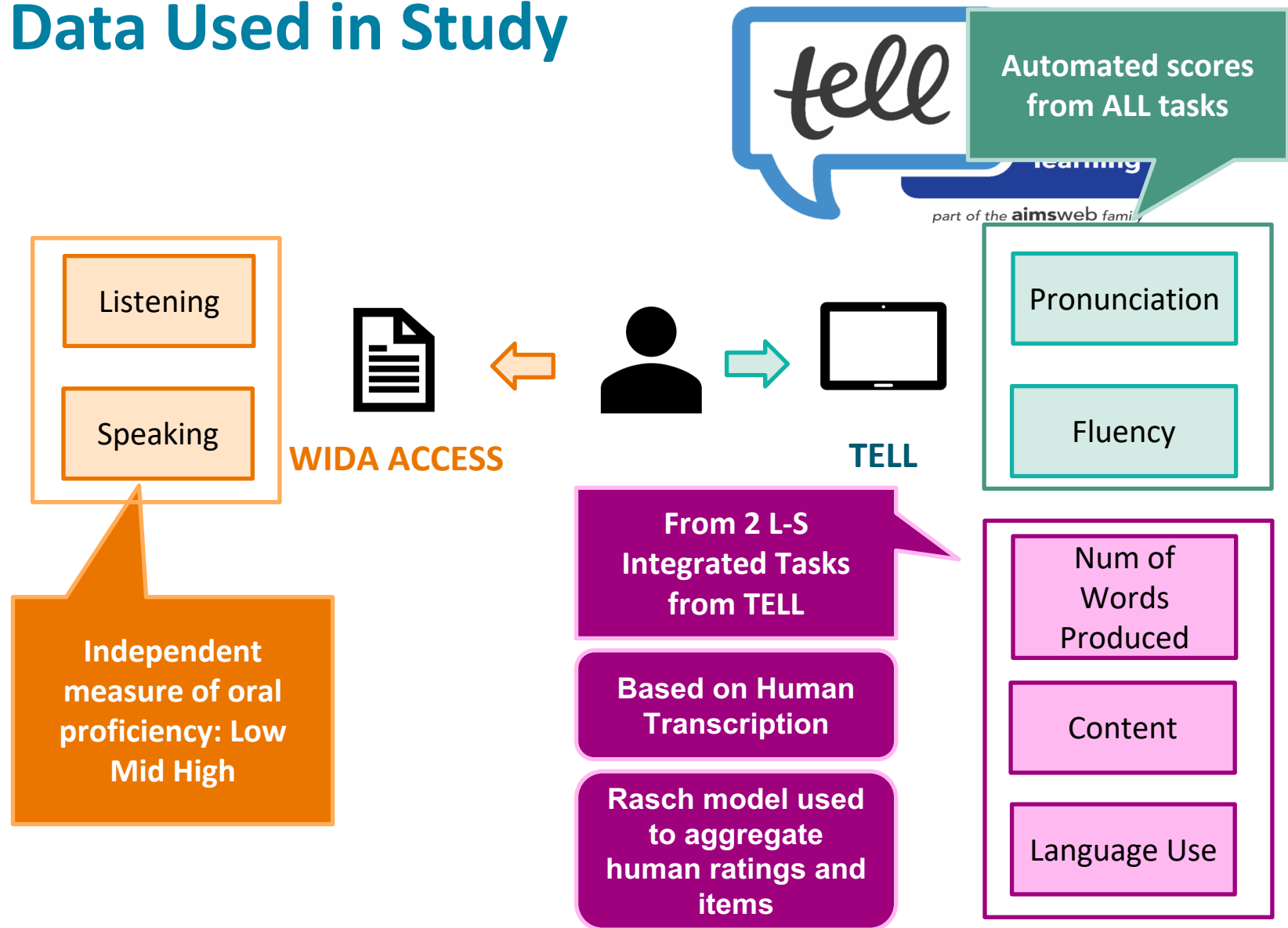


# Participants

- ELLs from 10 different states
- Majority of students speak Spanish as L1 (70%)

Grade band	<i>n</i>
1-2	181
3-5	270
6-8	166
9-12	159
Total	776

# Data Used in Study



# Integrated L-S Tasks

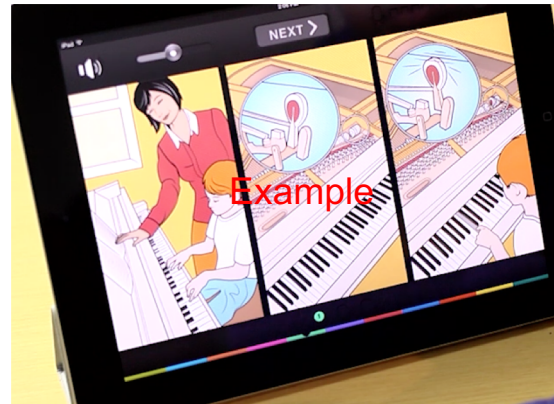
Three images  
presented one  
at a time

Descriptive Prompt

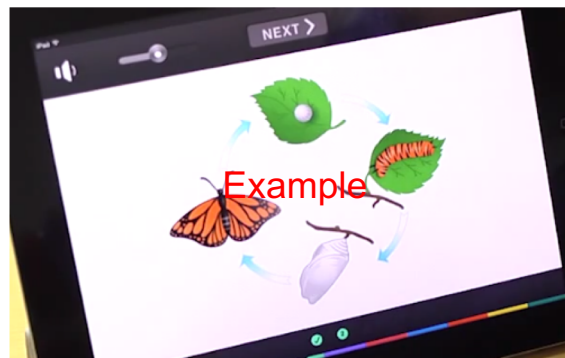
Expository Prompt

Task 1

Video



Task 2



# Integrated L-S Tasks

	Narrative Prompt	Expository Prompt
Task 1	<div>Grades 1-2 3 items</div>	
Task 2	<div>Grades 3-5 Grades 6-8 Grades 9-12  6 items</div>	

# Content & Language Use Rubric

- Scale of 0 to 3
- **Content**: the quality of the content (i.e., how well the important ideas from the text were represented)
- **Language Use**: the effective and accurate use of English language including the coherence of the response (c.f., *selection and sophistication of vocabulary usage was not included*)



# Three Analyses

## **1. Amount of Content**

Repeated measures ANOVA: Words produced by prompt type (within-subject) and oral proficiency level (between-subject) by grade band

## **2. Quality of Content and Language Use**

Repeated measures ANOVA: Human ratings by trait (within-subject), prompt type (within-subject) and oral proficiency level (between-subject) by grade band

## **3. Relationship of L2 Proficiency measures**

Structural equation model including all measures plus fluency and pronunciation measures by grade band

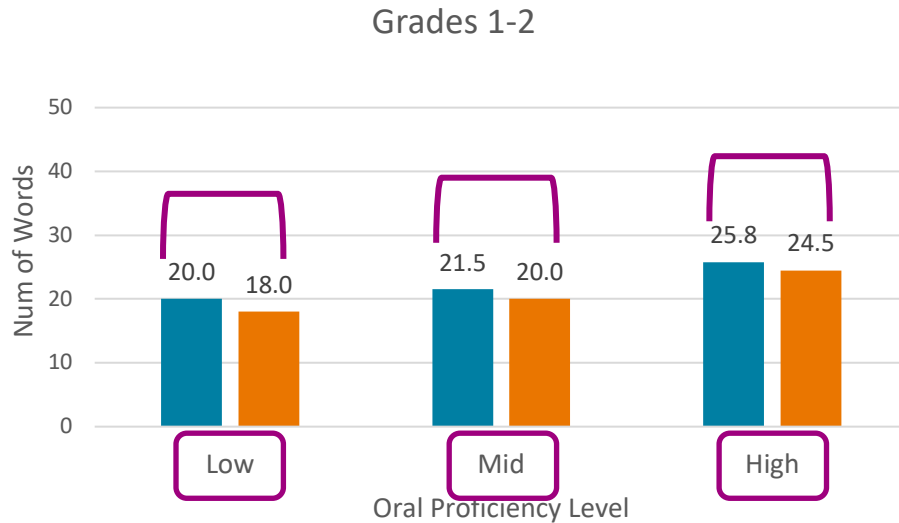
# Amount of Content

1. Do ELs at varying levels of L2 oral skills represent different **amounts of content** in their re-telling tasks?



# Amount of Content

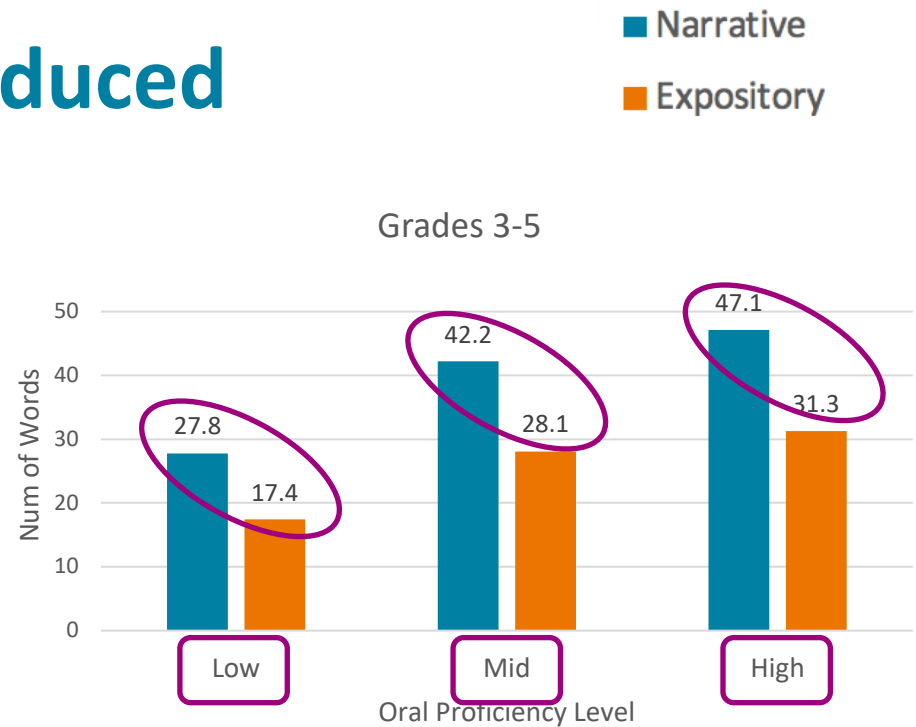
■ Narrative  
■ Expository



- The higher the oral proficiency level, the more the content produced ( $\eta_p^2 = .066$ )
- No difference between prompt types
- No difference between prompt type among oral proficiency

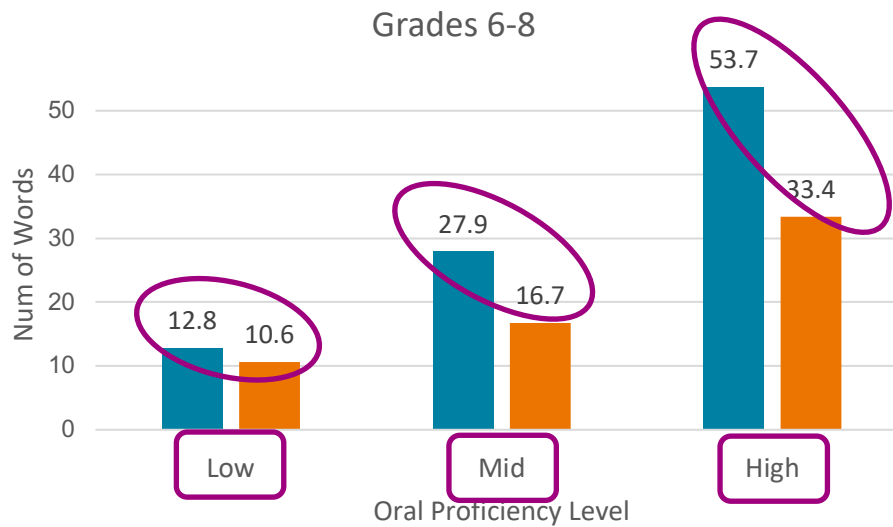
# Number of Words Produced

- The higher the oral proficiency level, the more the content produced ( $\eta_p^2 = .179$ )
- More words produced from Narrative prompt than from Expository prompts ( $\eta_p^2 = .398$ )
- No difference between prompt type among oral proficiency

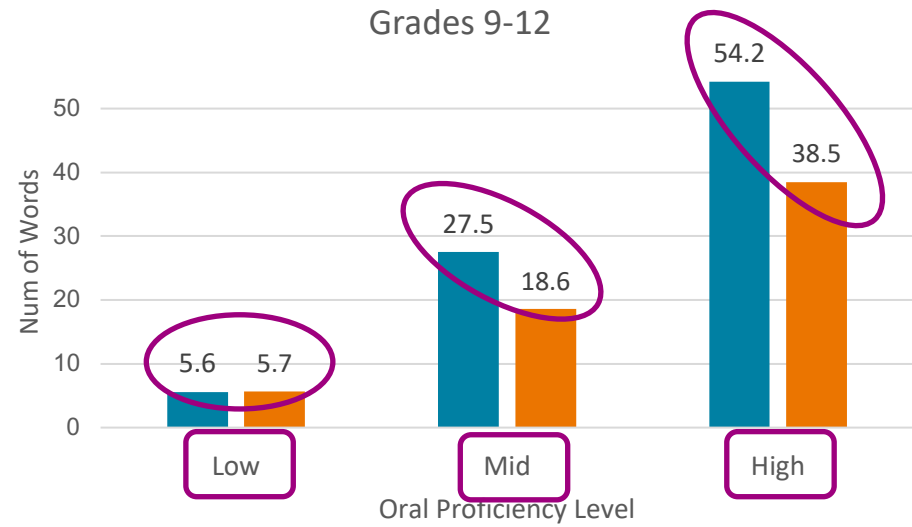


# Number of Words Produced

■ Narrative  
■ Expository



- The higher the oral proficiency level, the more the content produced ( $\eta_p^2 = .371$ )
- More words produced from Narrative prompt than from Expository prompts ( $\eta_p^2 = .402$ )
- The difference between narrative and expository increases as the oral proficiency gets higher ( $\eta_p^2 = .223$ )



- The higher the oral proficiency level, the more the content produced ( $\eta_p^2 = .552$ )
- More words produced from Narrative prompt than from Expository prompts ( $\eta_p^2 = .215$ )
- The difference between narrative and expository increases as the oral proficiency gets higher ( $\eta_p^2 = .148$ )

# Amount of Content

- Students at higher levels of oral proficiency generally produced more content
  - The differences among the levels were quite small for grades 1-2 students, larger for grades 3-5, very large for grades 6-8 and grades 9-12
- Narratives yielded far more words than expository prompts at all grades except 1-2

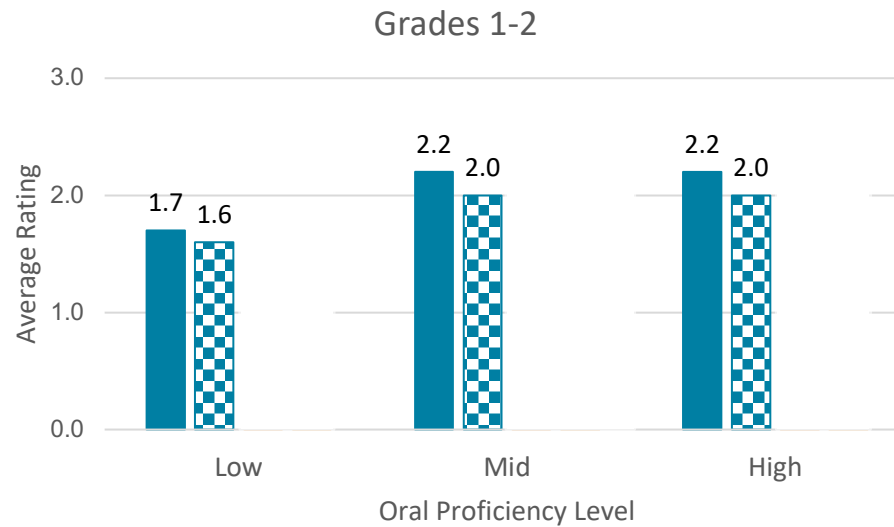
# Quality of Content and Language Use

2. Do ELs of varying levels of L2 oral proficiency demonstrate difference in levels of performance on **content** and **language use**?



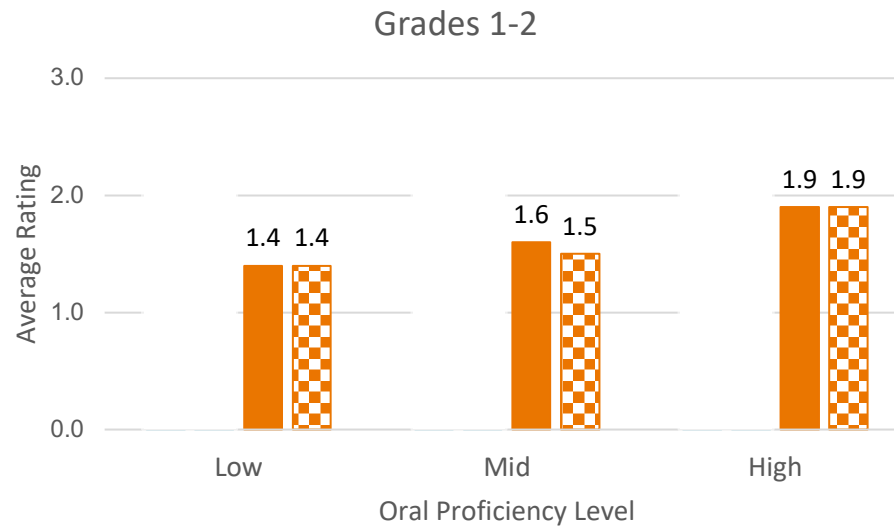
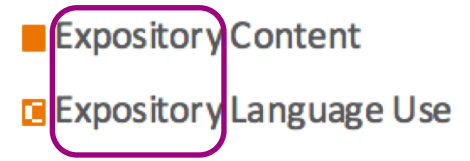
# Quality of Content and Language Use

- Narrative Content
- Narrative Language Use

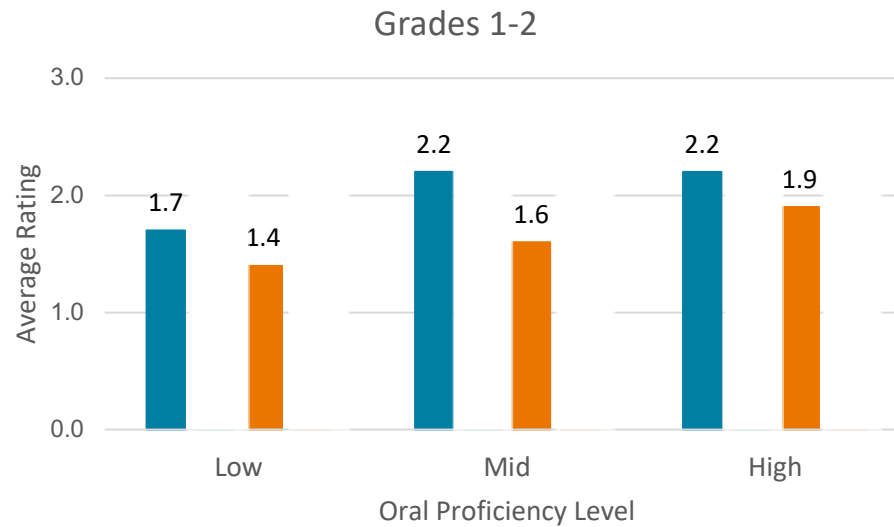
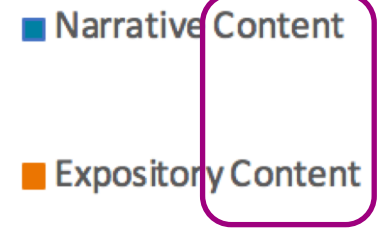




# Quality of Content and Language Use

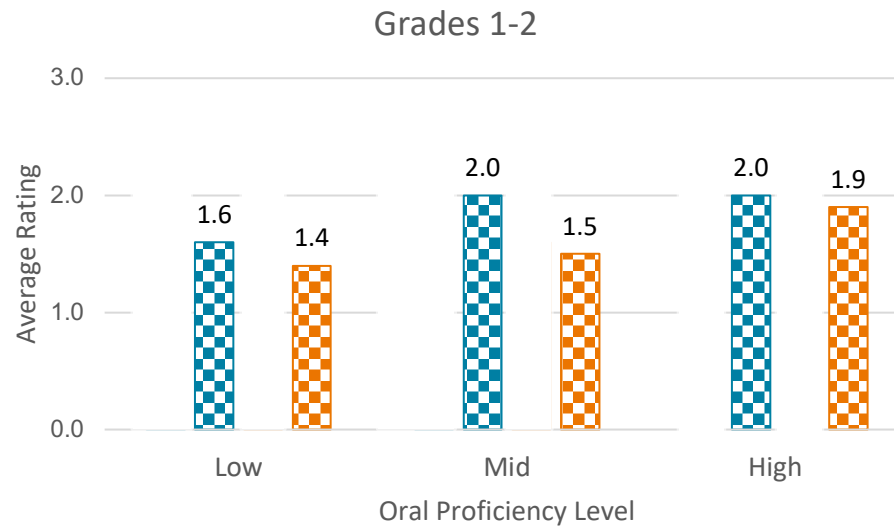


# Quality of Content and Language Use



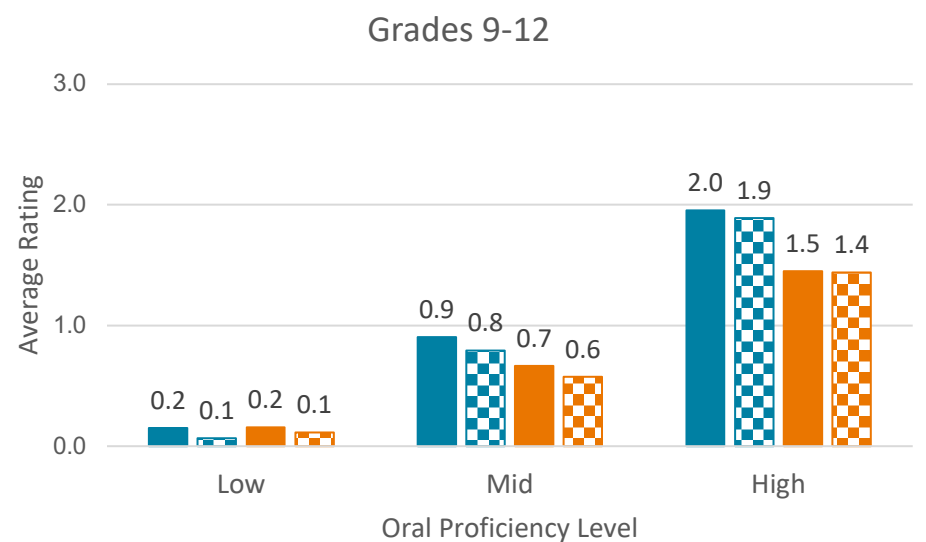
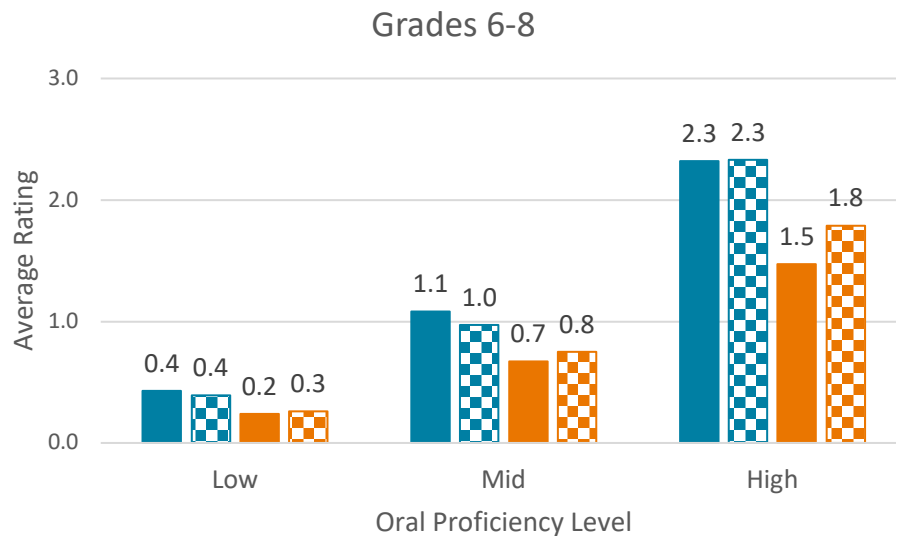
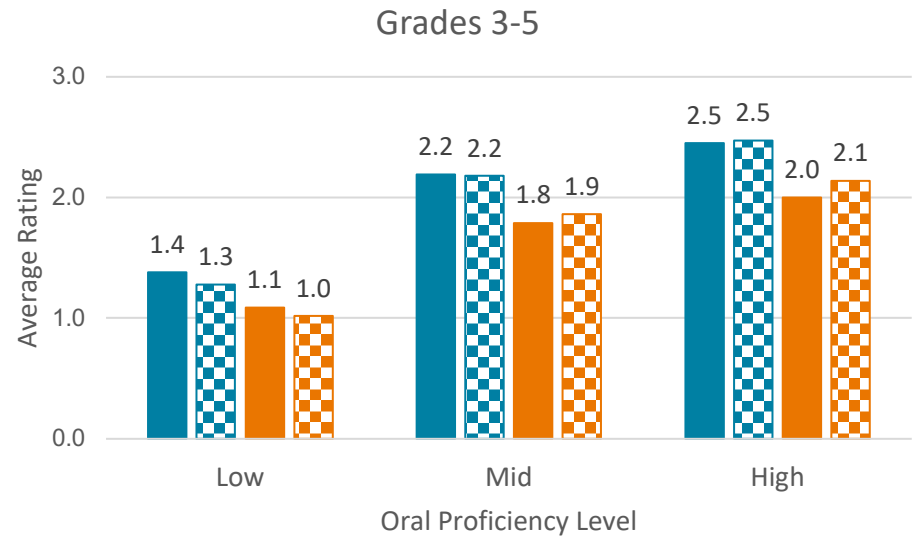
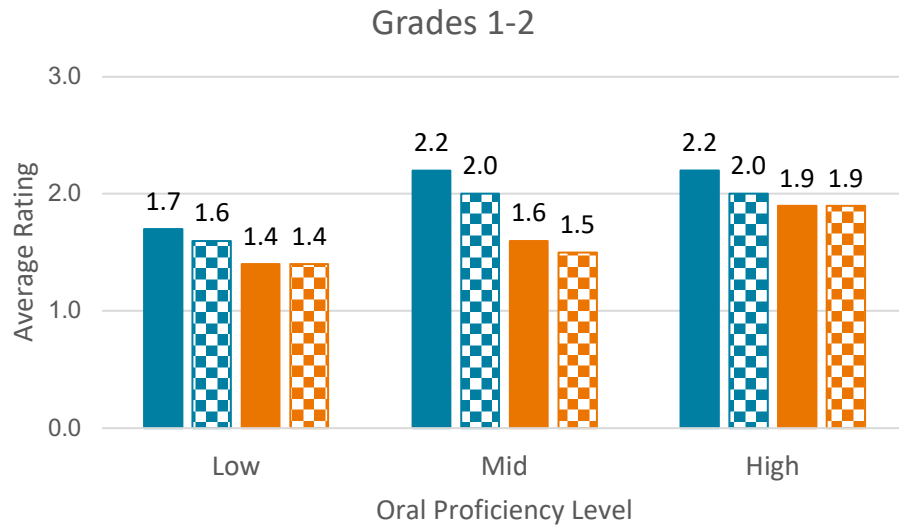
# Quality of Content and Language Use

- Narrative Language Use
- Expository Language Use

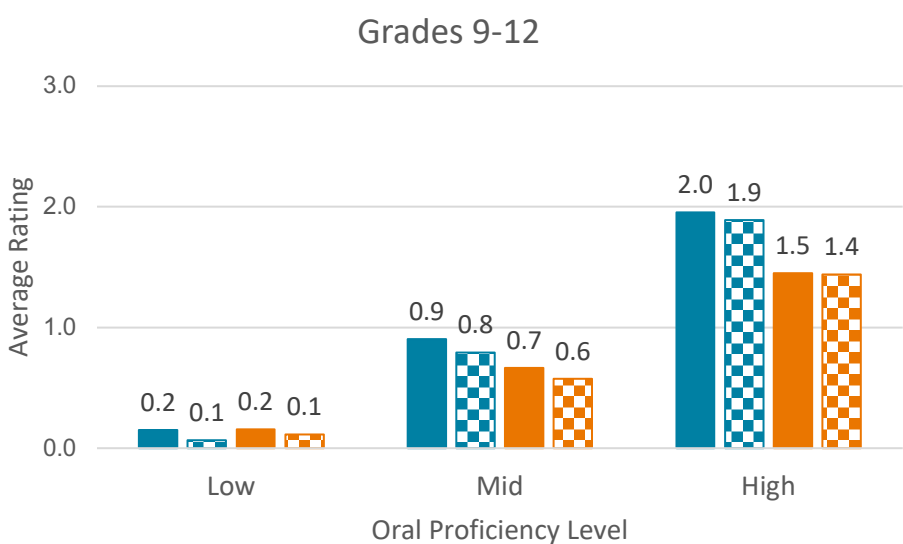
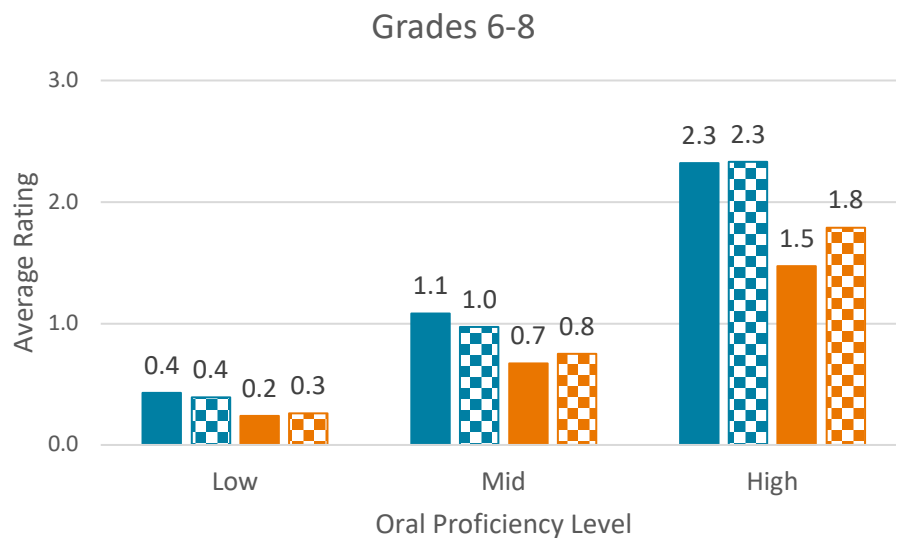
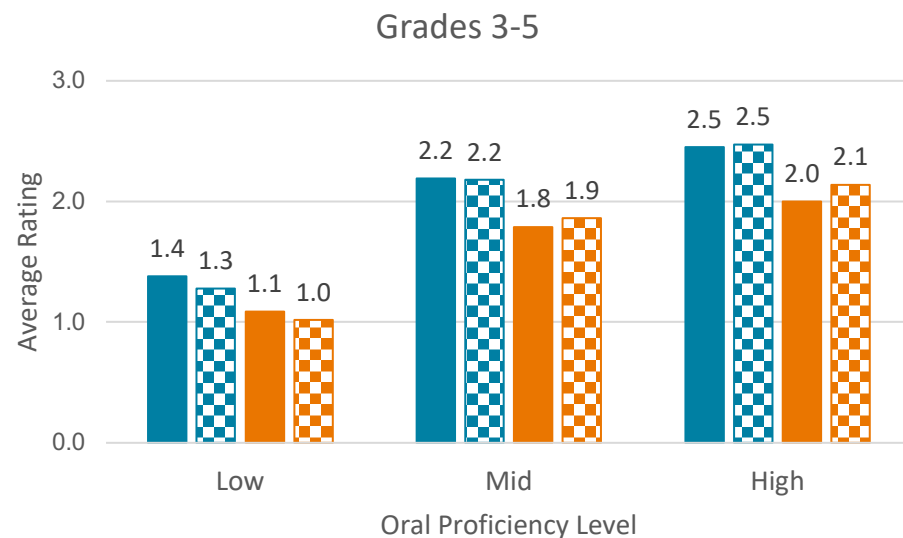
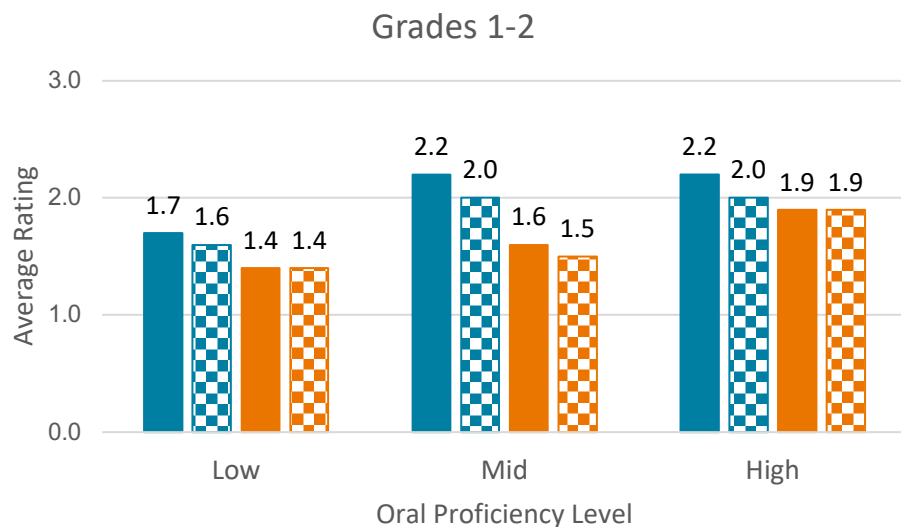


# Quality of Content and Language Use

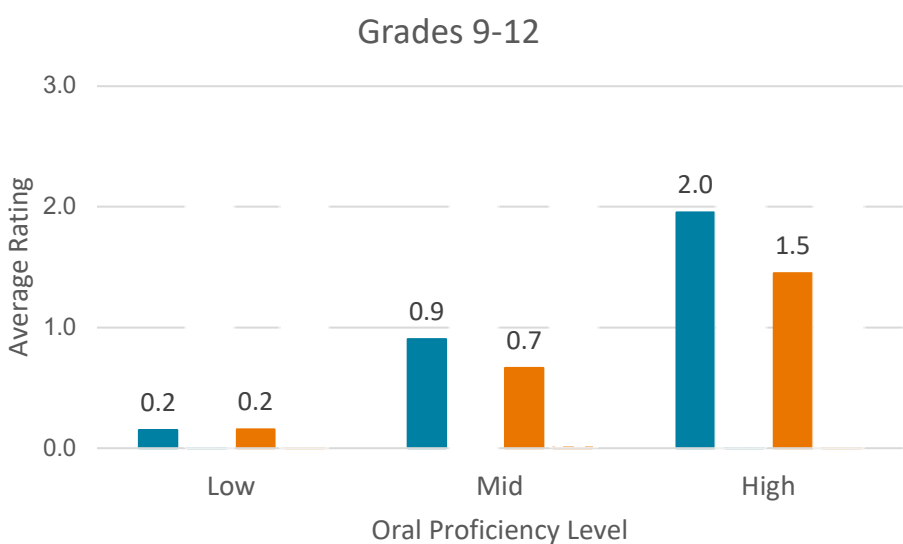
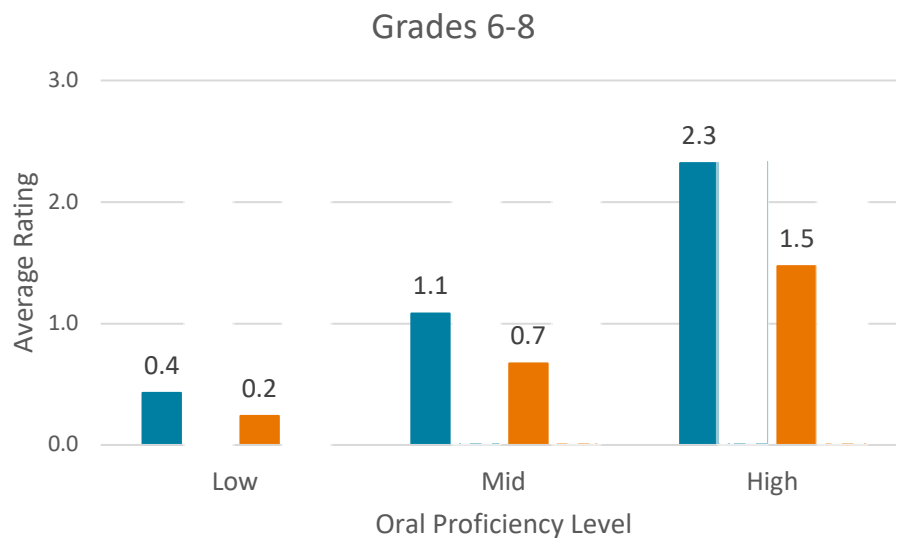
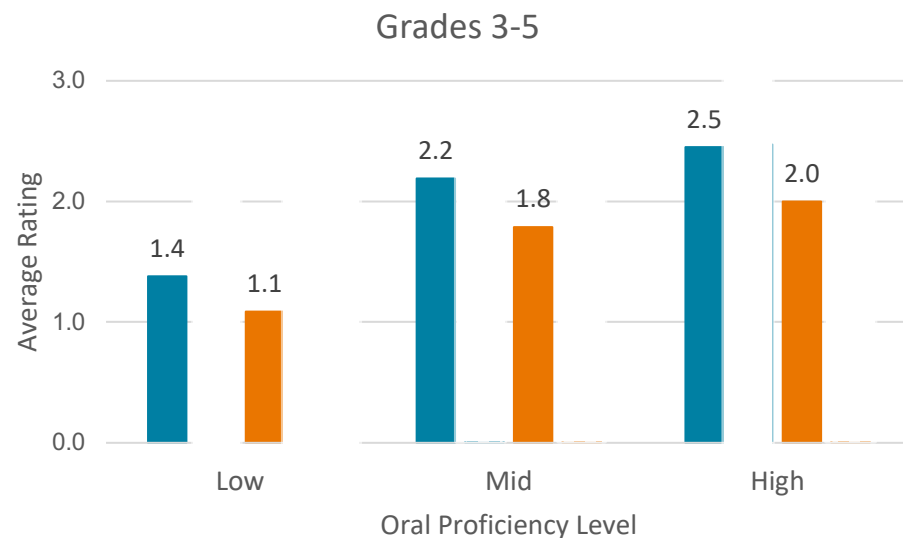
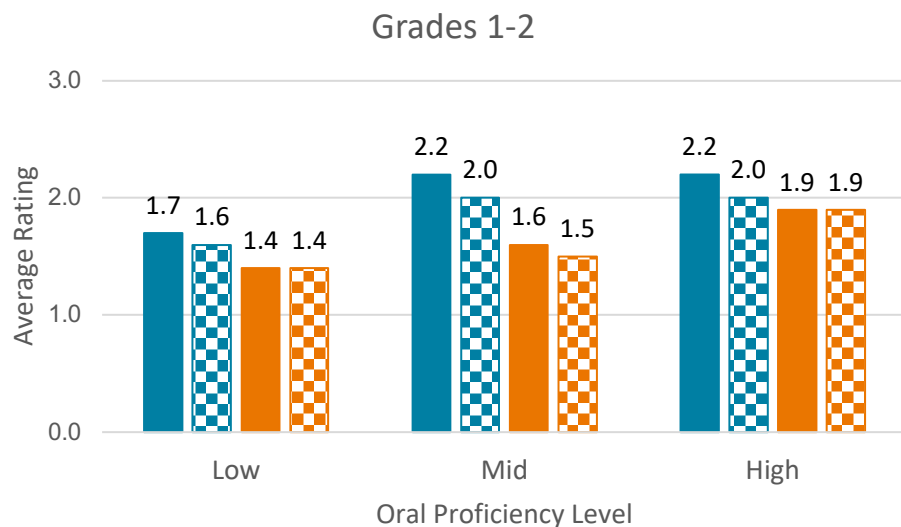
- Narrative Content
- ▣ Narrative Language Use
- Expository Content
- ▣ Expository Language Use



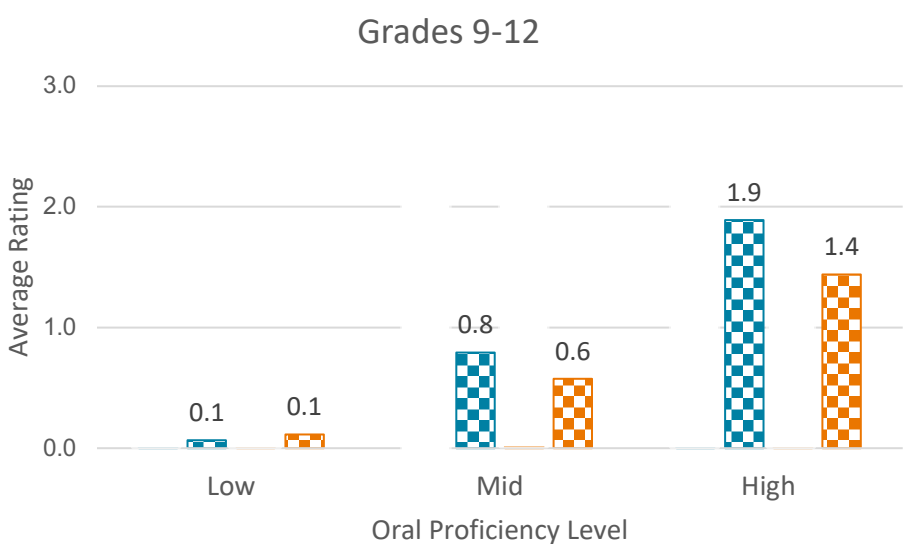
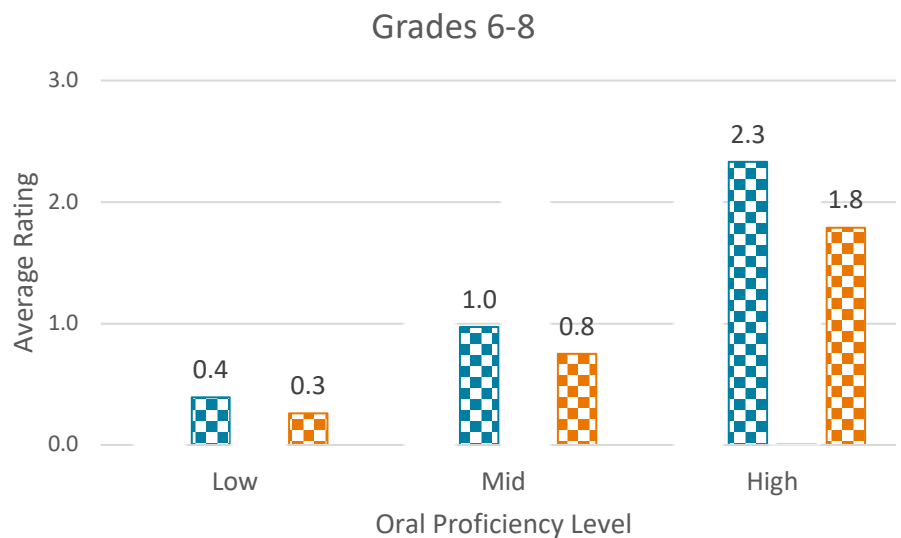
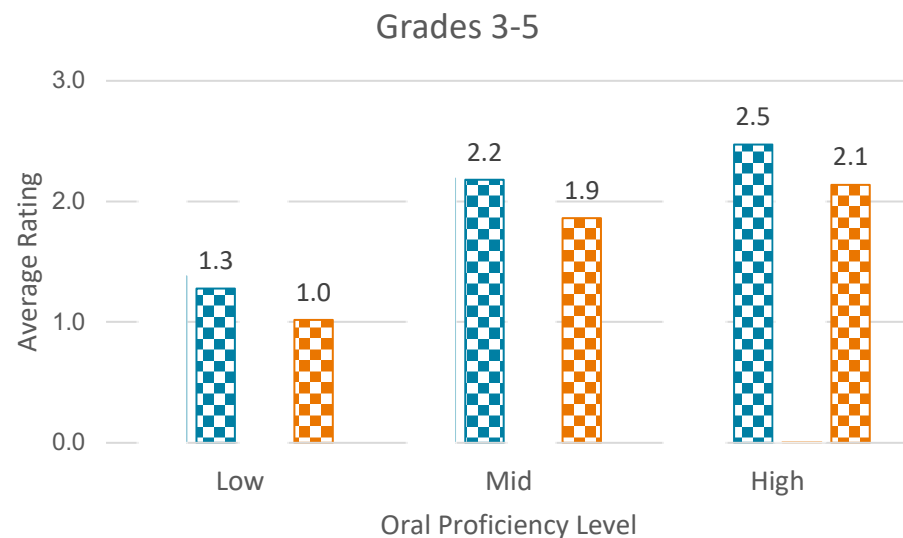
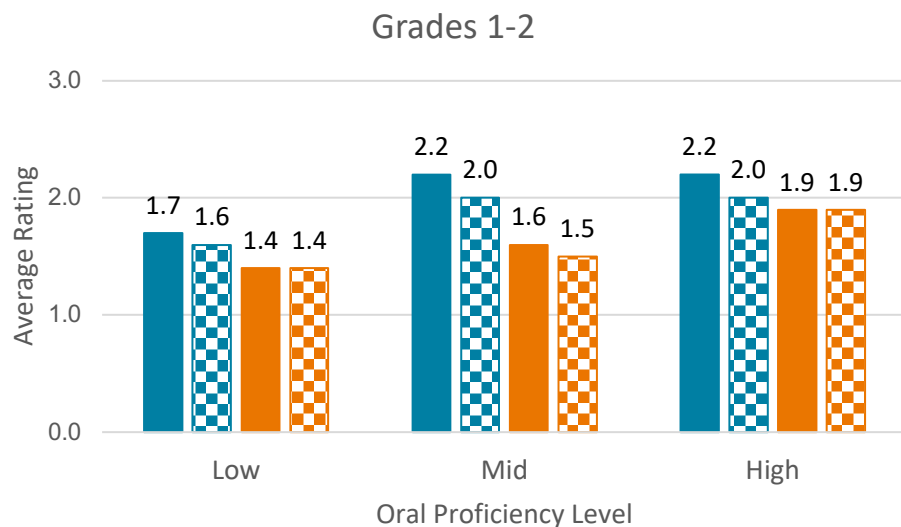
Across all grade levels higher than 1-2, the higher the proficiency level, the better ratings were achieved for both content and language use



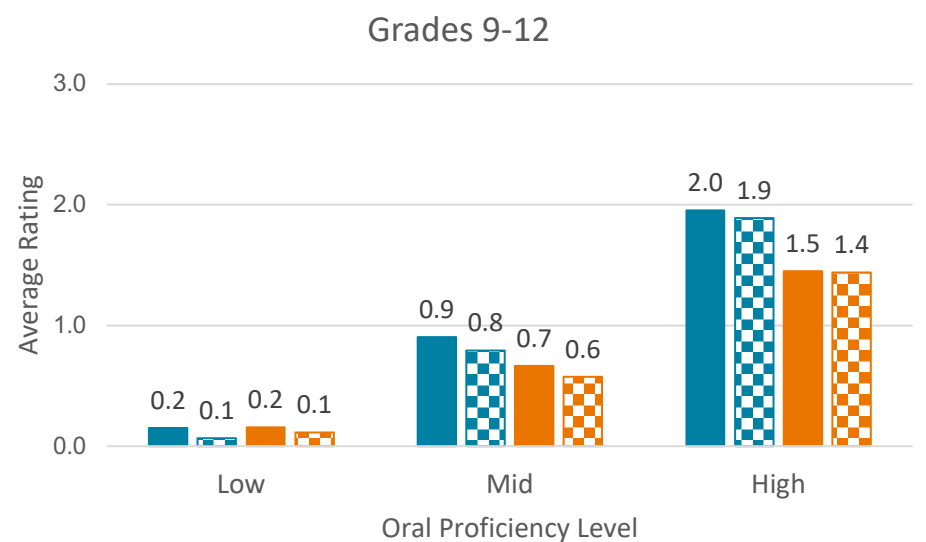
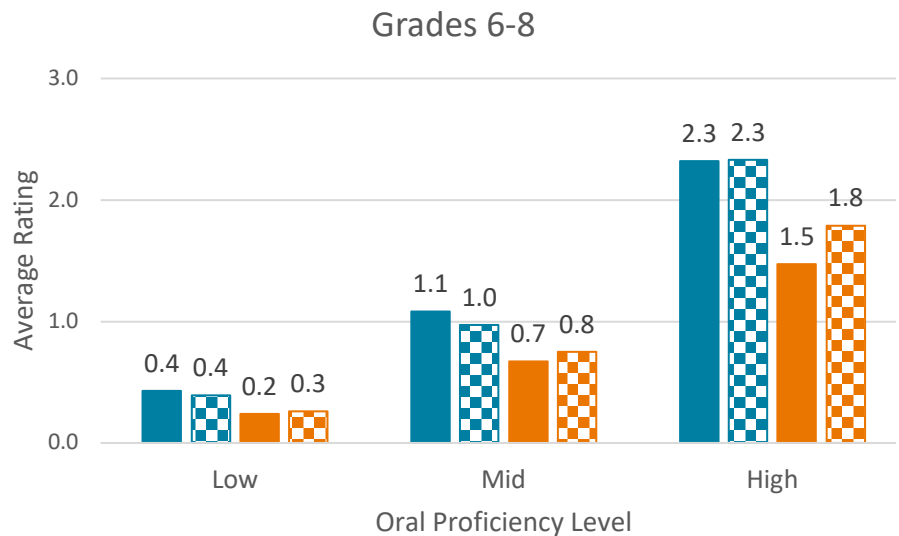
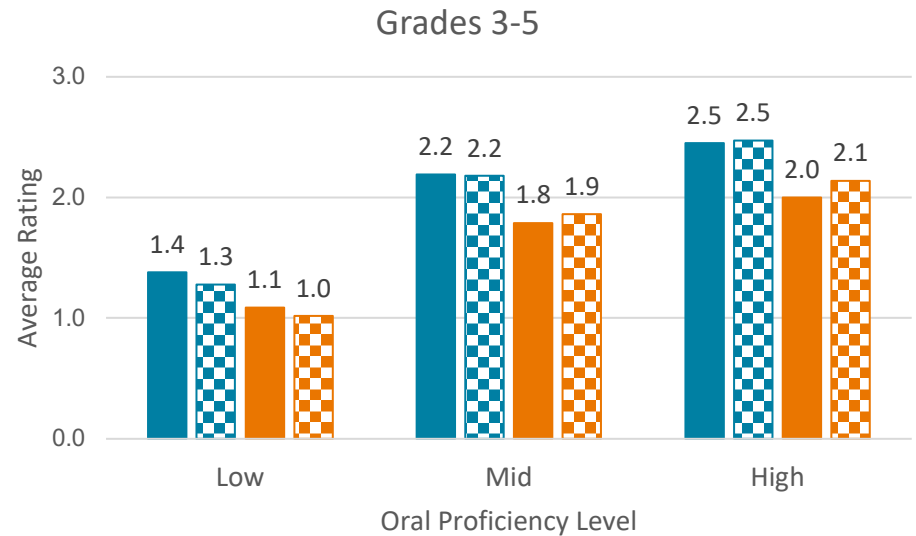
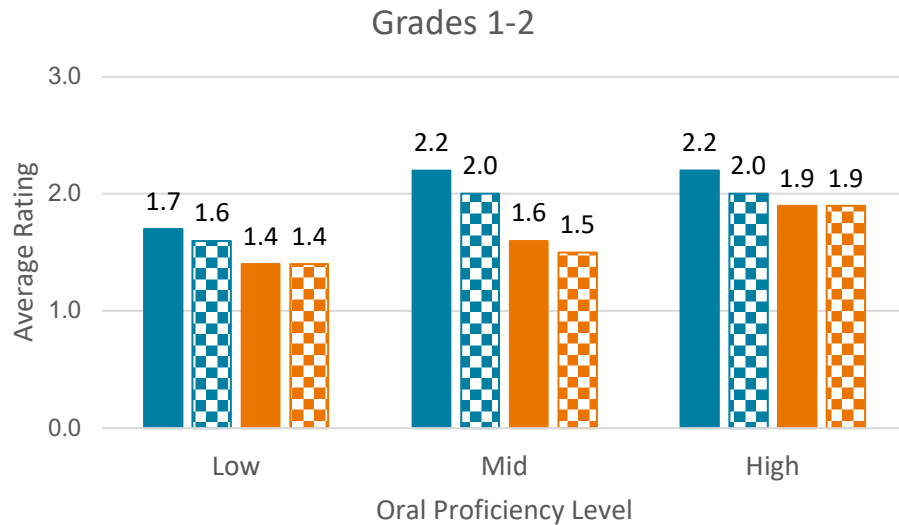
Across all grade levels higher than 1-2, the higher the proficiency level, the better ratings were achieved for both content and language use



Across all grade levels higher than 1-2, the higher the proficiency level, the better ratings were achieved for both content and language use

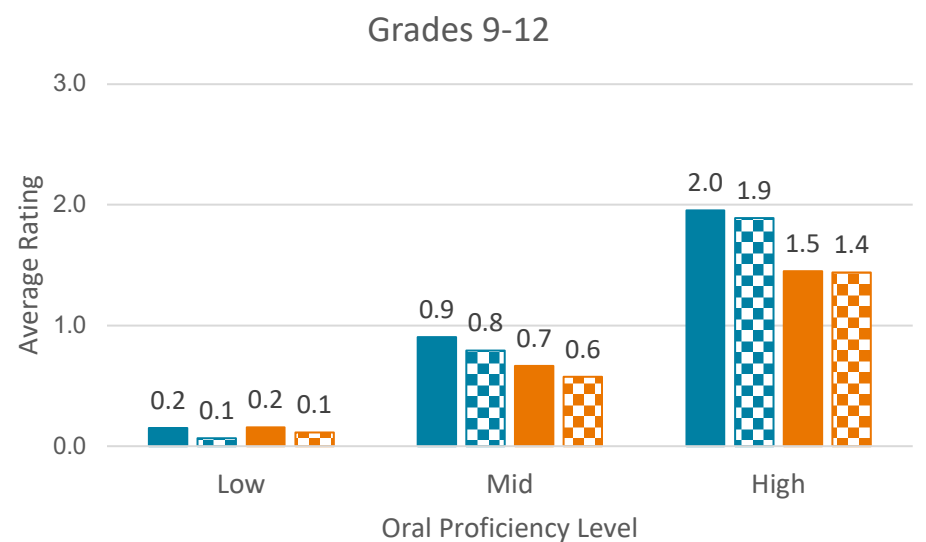
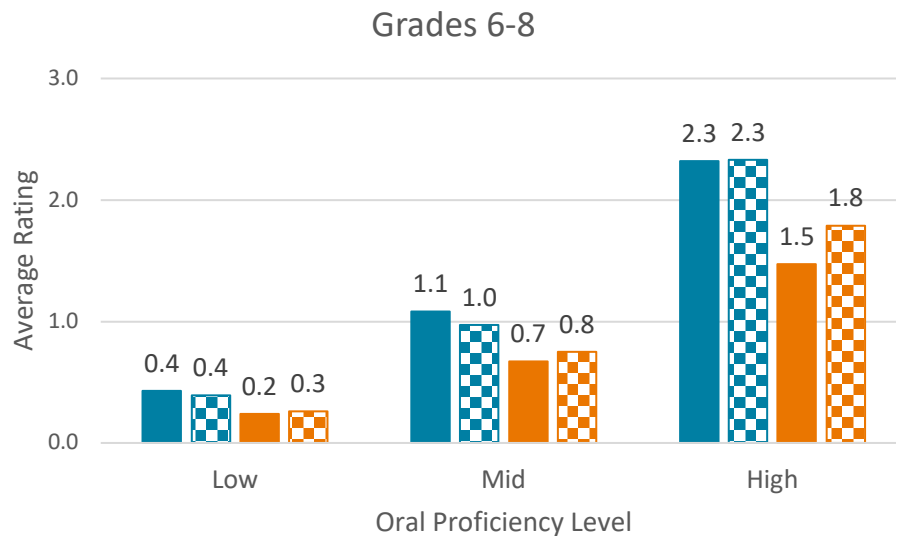
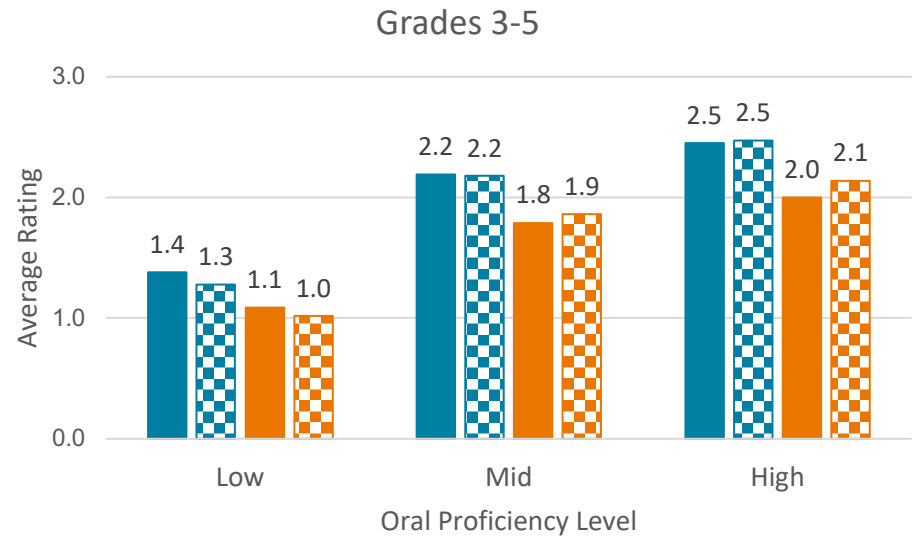
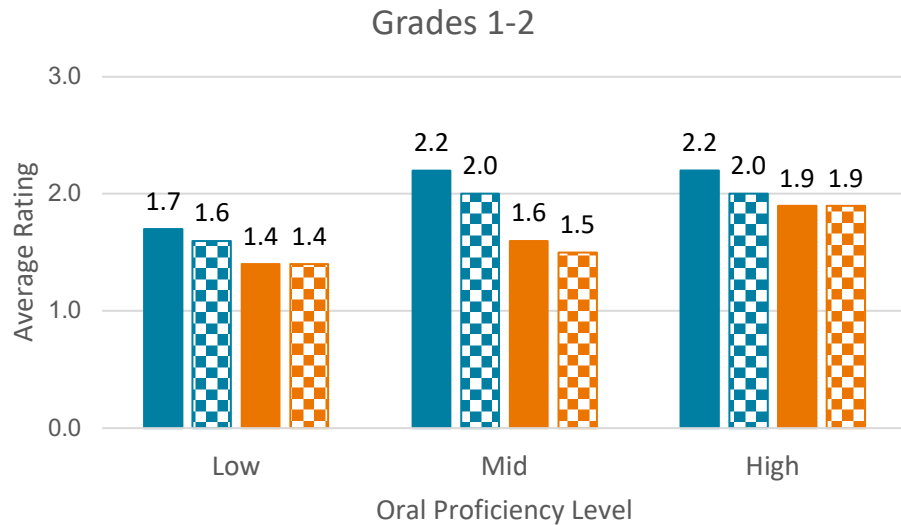


Effect even stronger at higher grade bands  
- Increased task demands and complexity of the prompts

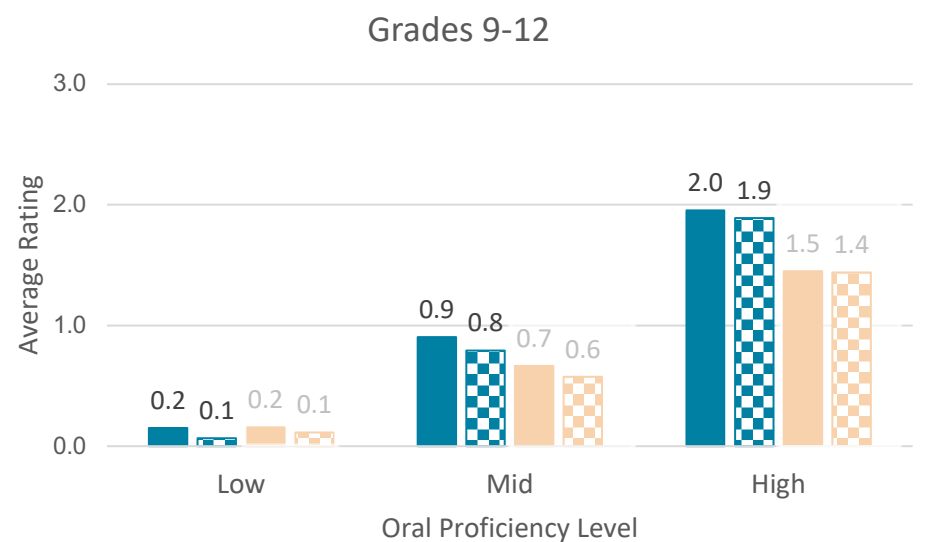
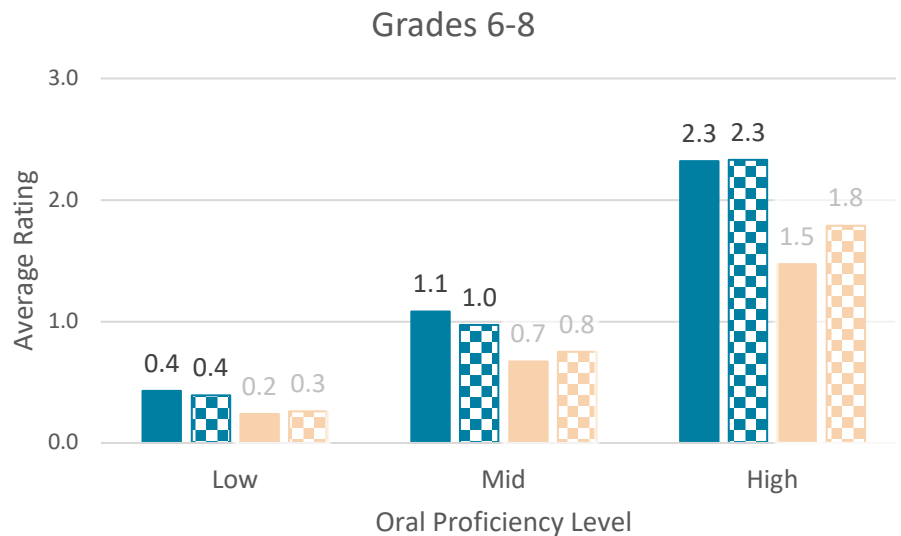
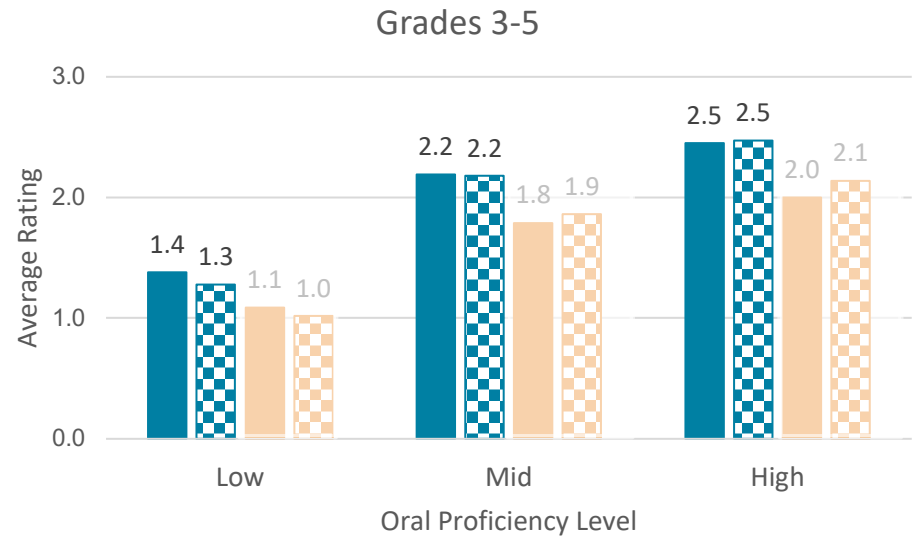
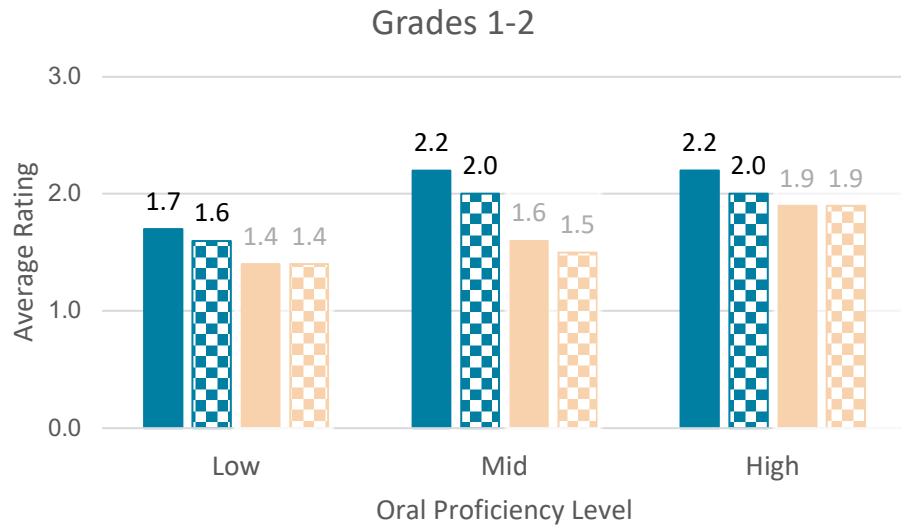




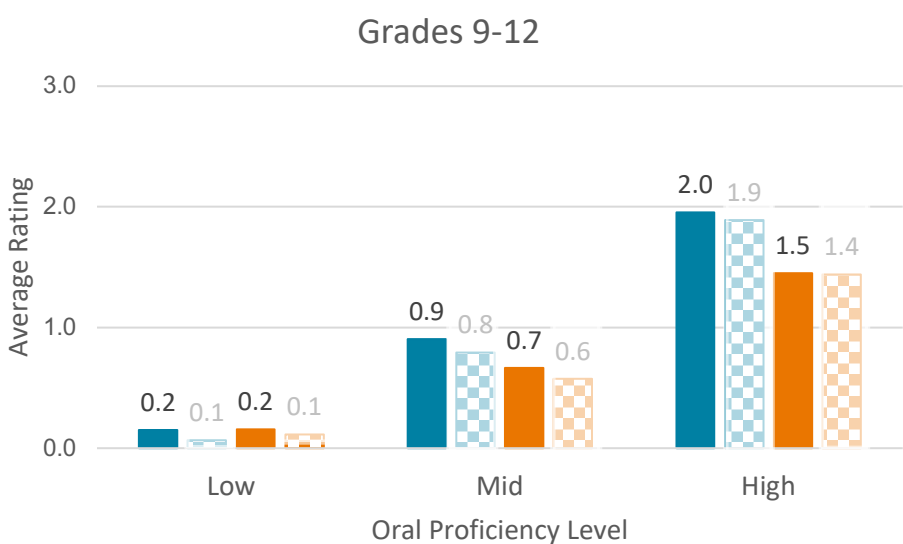
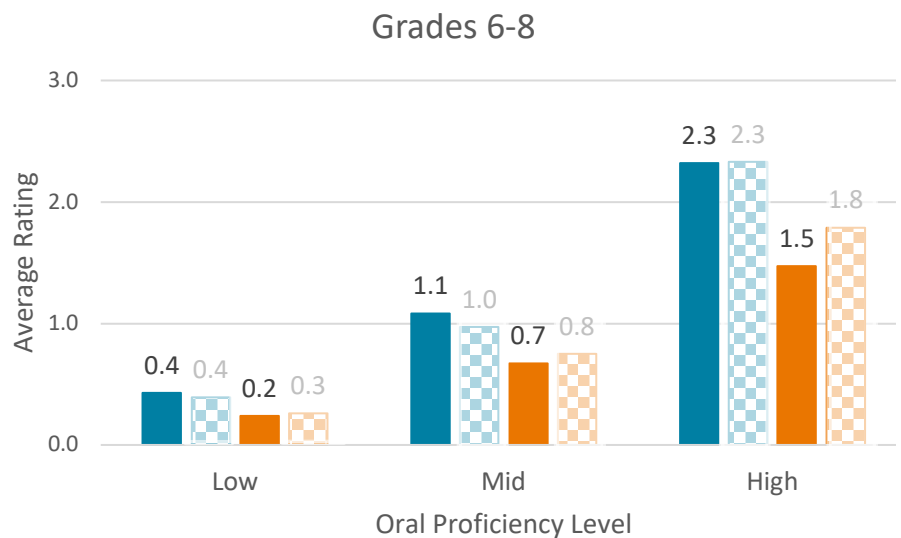
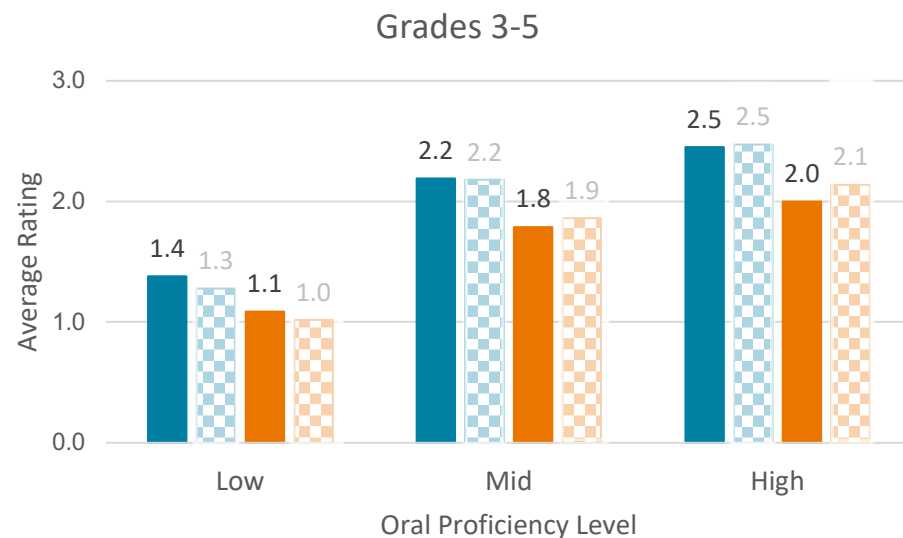
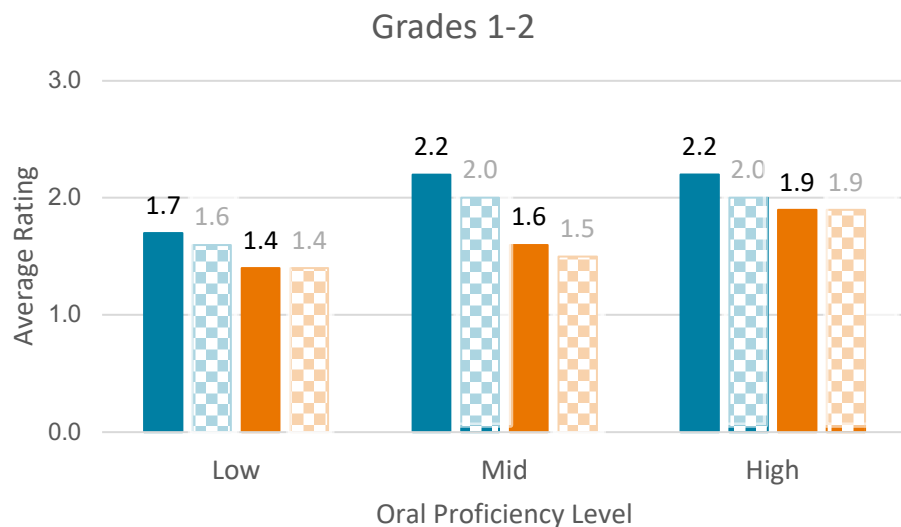
Narratives produce higher ratings on both content and language use than expository prompts



Narratives produce higher ratings on both content and language use than expository prompts



Trait (Content and Language Use) is a very consistent effect of content scores being higher than language use



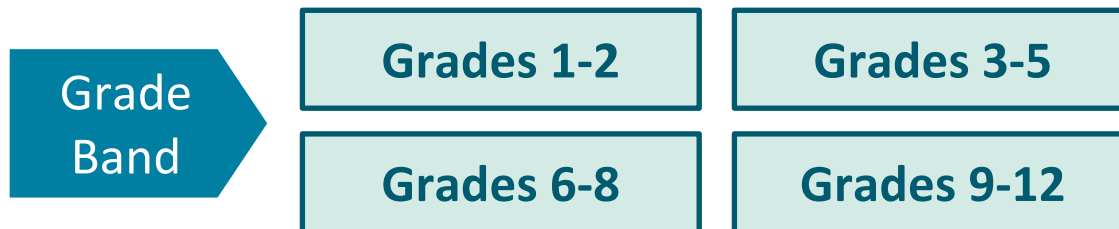
# Quality of Content and Language Use

- Generally, students at high levels of oral proficiency received a higher score on content and language use
  - Esp. for the upper grade bands
- Narrative prompts produced higher ratings than expository prompts on content and language use
- Content scores were higher than language use scores

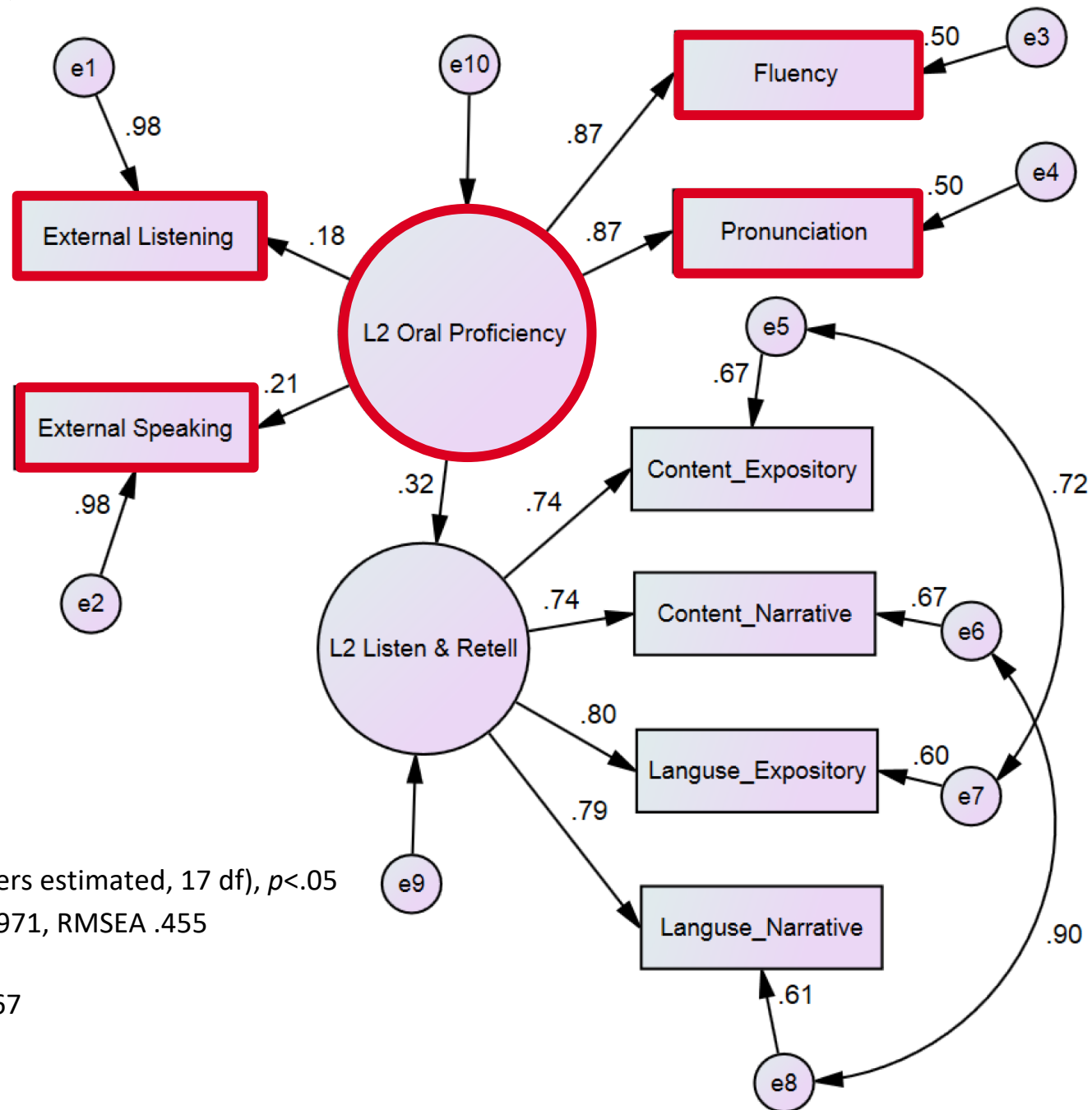
# Relationship of L2 proficiency measures

3-a. What do the measurement models tell us about the variables and their relationships?

3-b. What does the structural portion of the model tell us about the relationship between L2 Oral Proficiency and L2 Listen & Retell?



# Grades 1-2



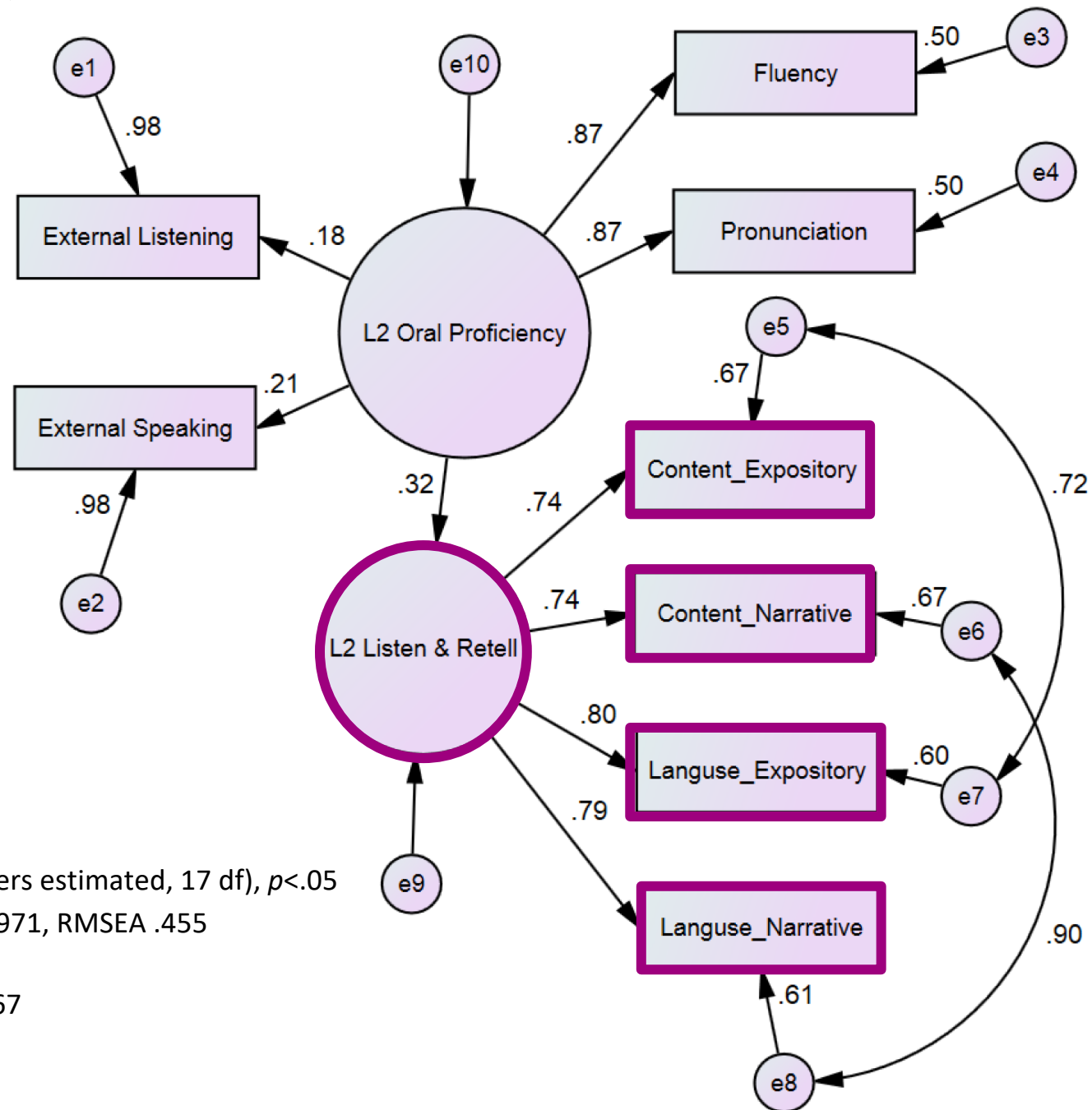
Chi-square = 66.5 (19 parameters estimated, 17 df),  $p < .05$

**Independence model** AIC 646.971, RMSEA .455

**Saturated model** AIC 72.0

**Model** AIC 104.533, RMSEA .167

# Grades 1-2



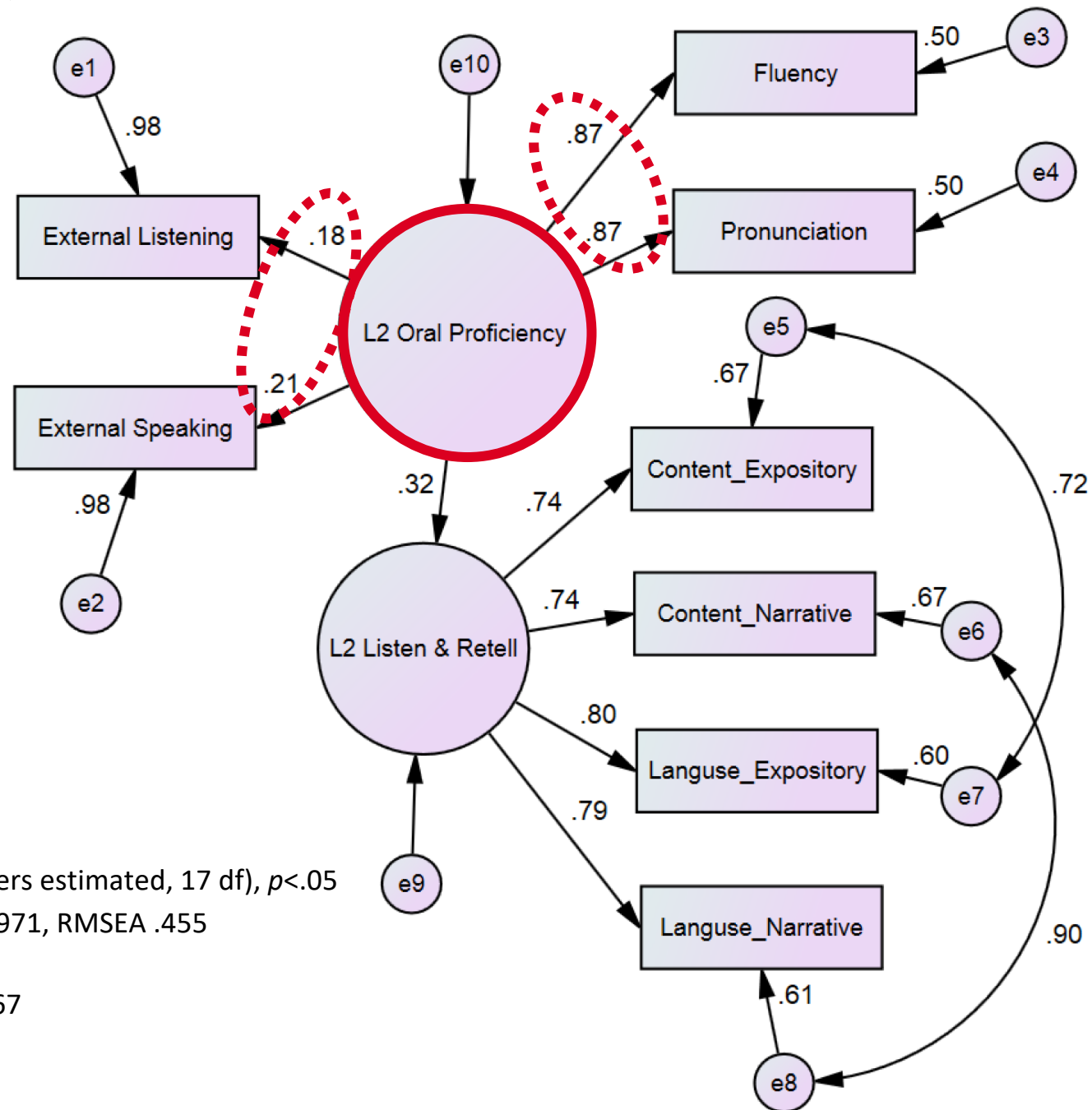
Chi-square = 66.5 (19 parameters estimated, 17 df),  $p < .05$

**Independence model** AIC 646.971, RMSEA .455

**Saturated model** AIC 72.0

**Model** AIC 104.533, RMSEA .167

# Grades 1-2



Chi-square = 66.5 (19 parameters estimated, 17 df),  $p < .05$

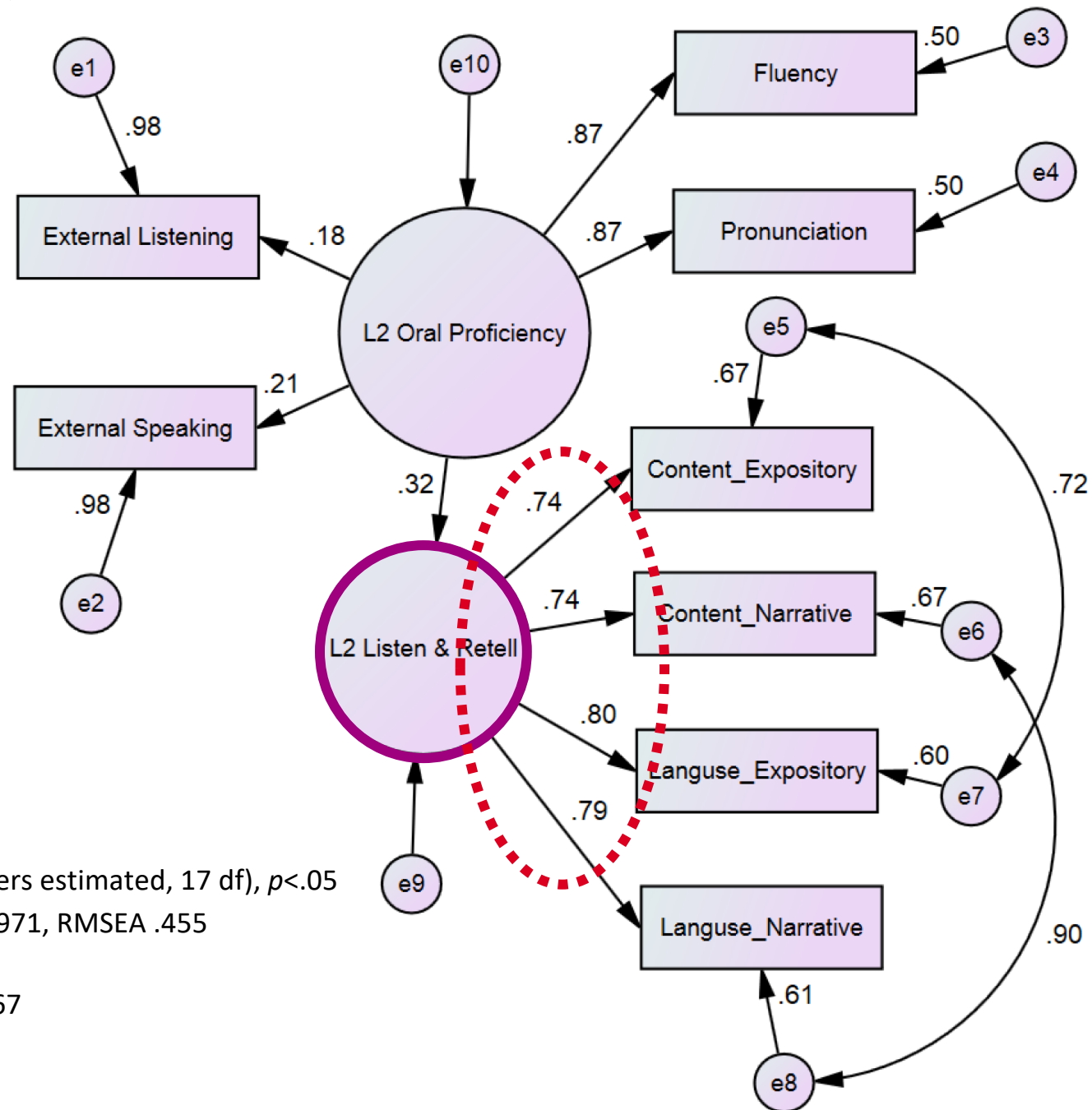
**Independence model** AIC 646.971, RMSEA .455

**Saturated model** AIC 72.0

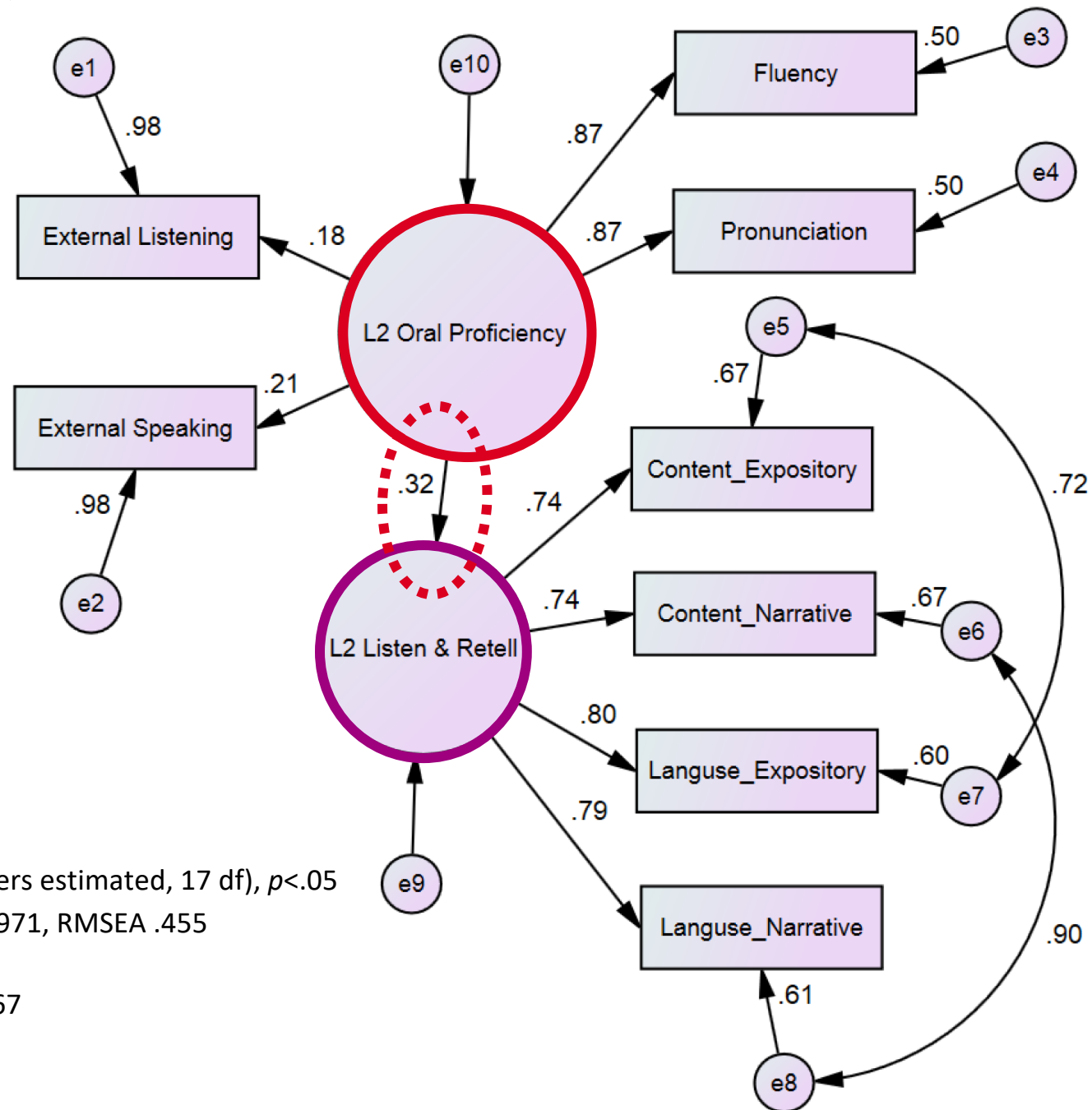
**Model** AIC 104.533, RMSEA .167



# Grades 1-2



# Grades 1-2



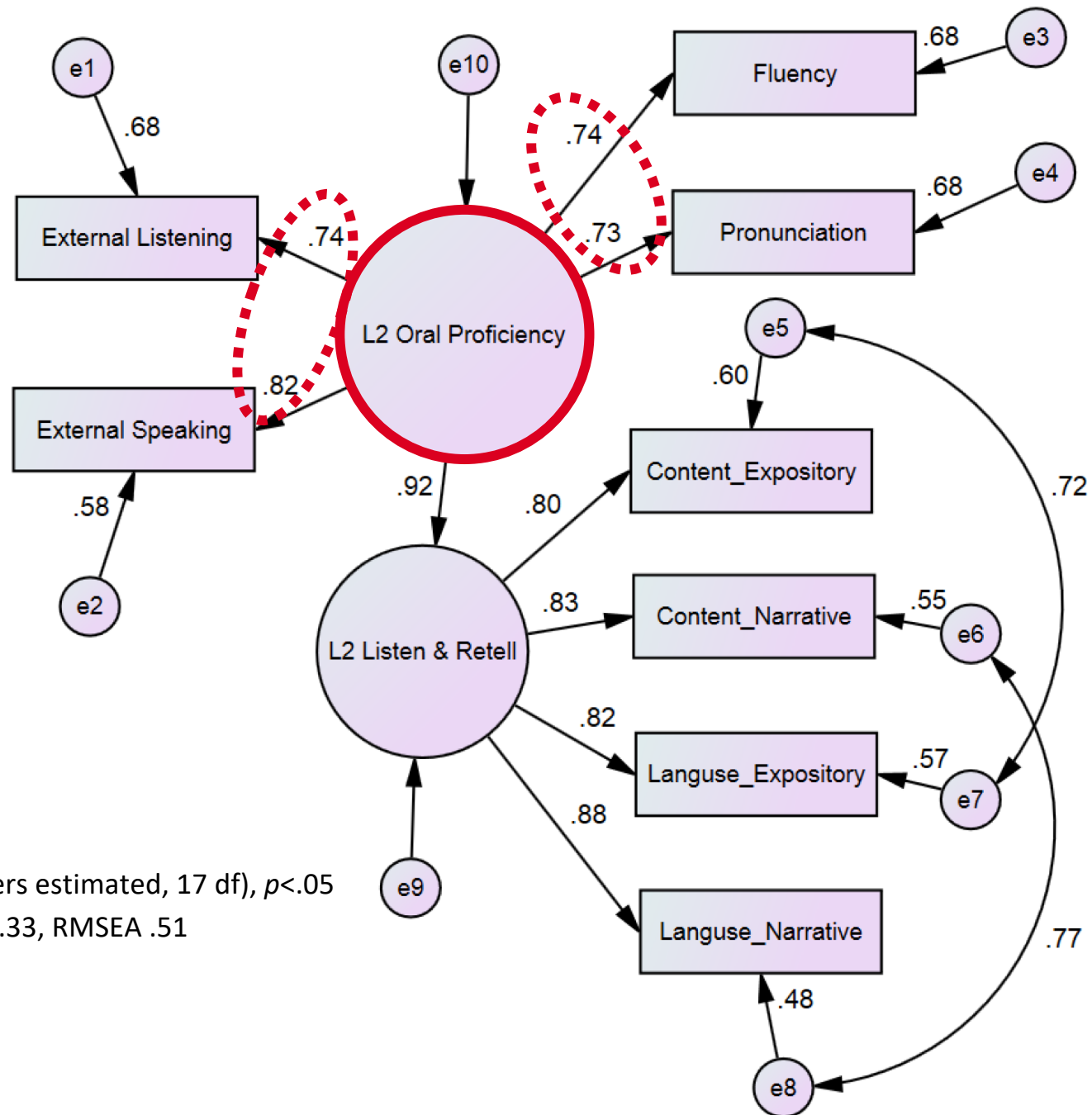
Chi-square = 66.5 (19 parameters estimated, 17 df),  $p < .05$

**Independence model** AIC 646.971, RMSEA .455

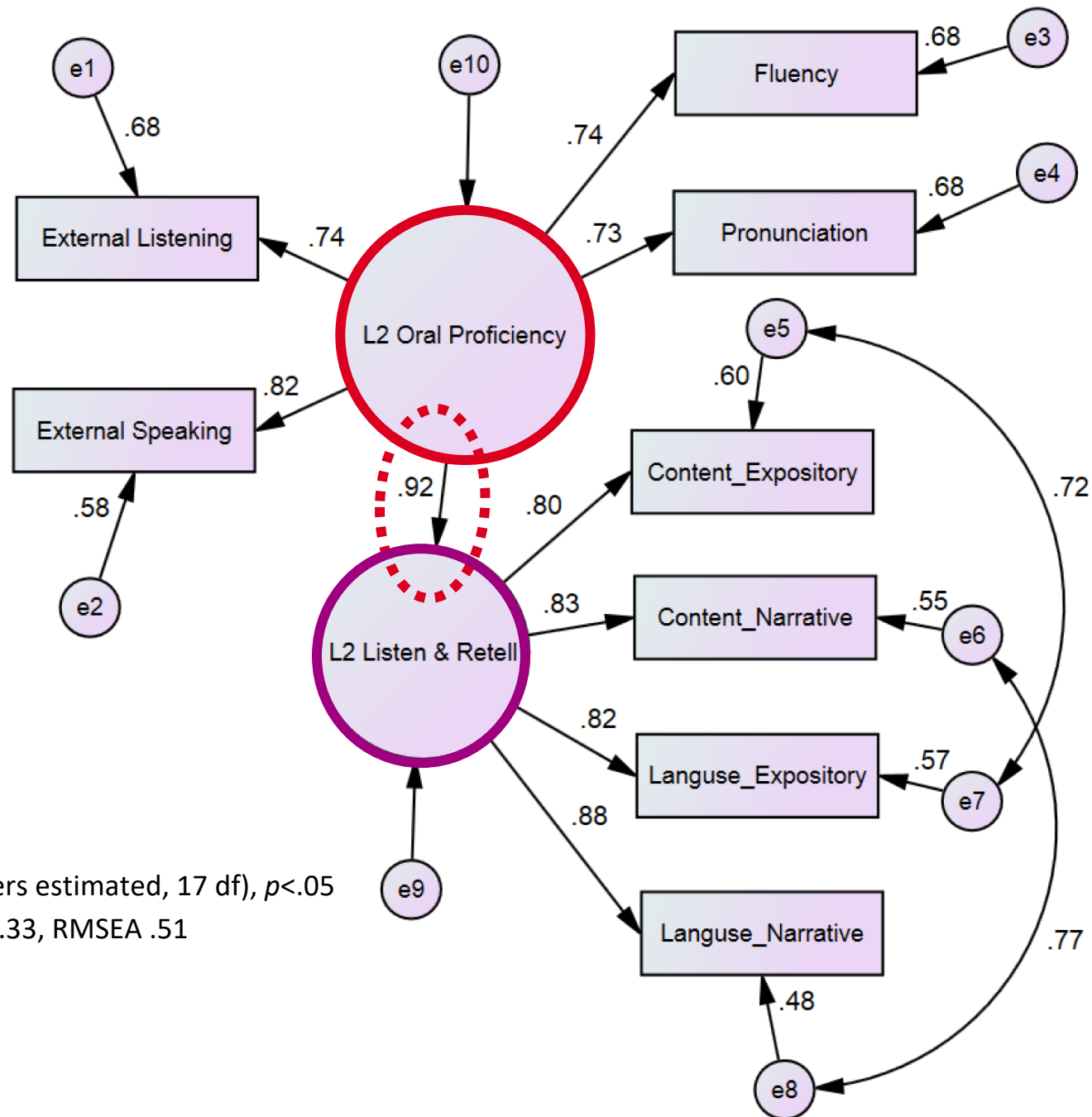
**Saturated model** AIC 72.0

**Model** AIC 104.533, RMSEA .167

# Grades 3-5



# Grades 3-5



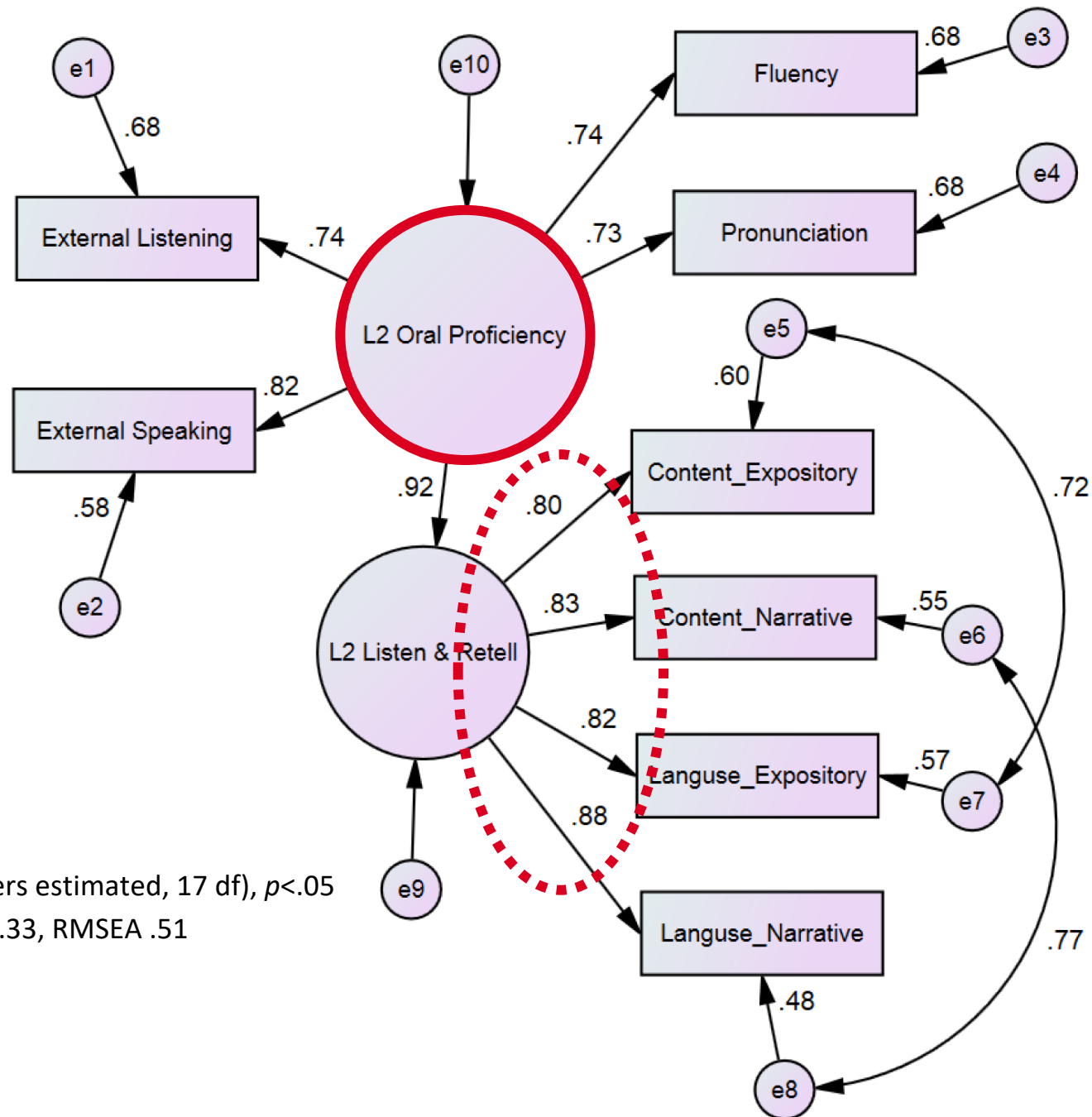
Chi-square = 89.6 (19 parameters estimated, 17 df),  $p < .05$

**Independence model** AIC 1740.33, RMSEA .51

**Saturated model** AIC 72.0

**Model** AIC 127.61, RMSEA .14

# Grades 3-5



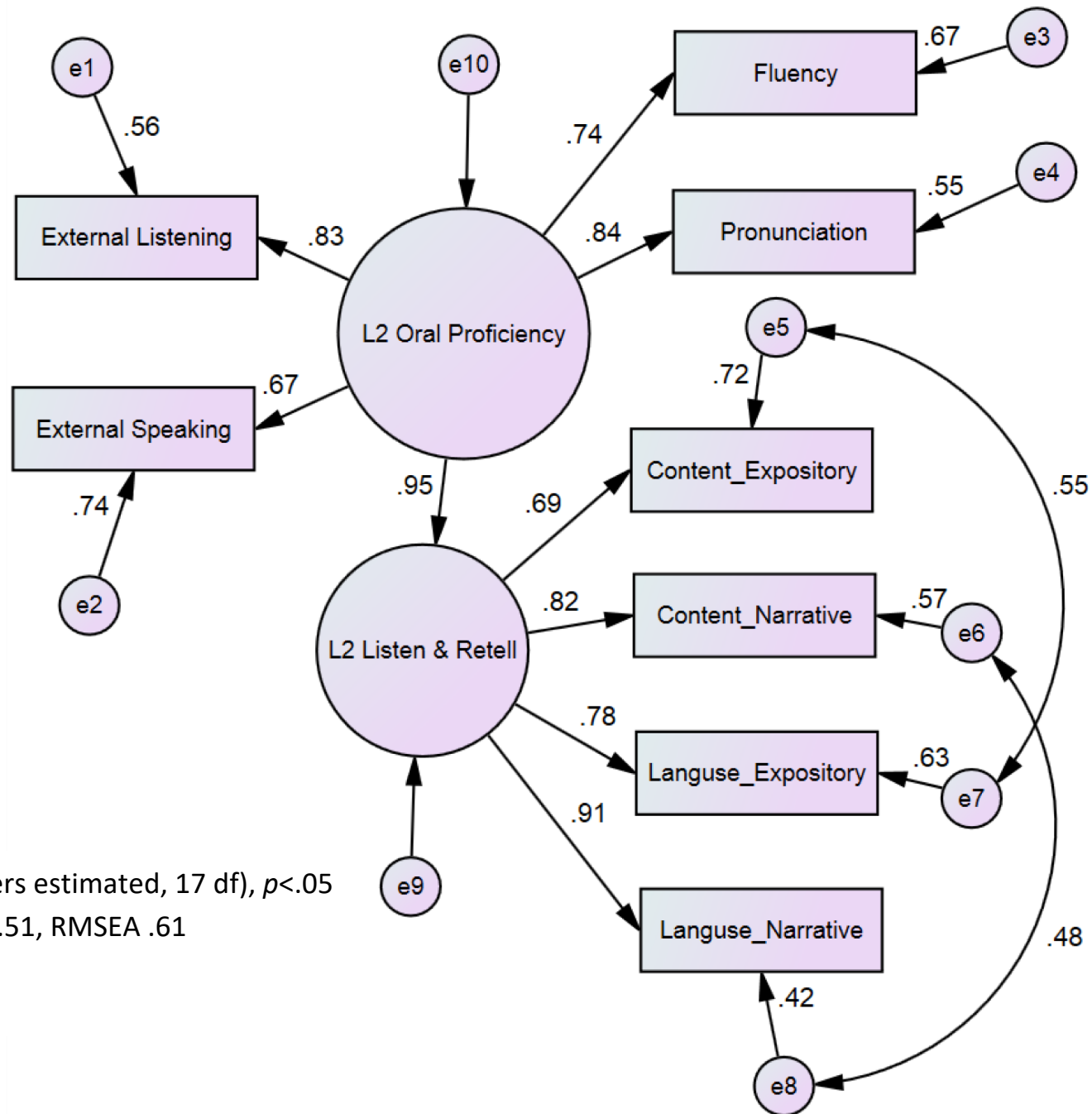
Chi-square = 89.6 (19 parameters estimated, 17 df),  $p < .05$

**Independence model** AIC 1740.33, RMSEA .51

**Saturated model** AIC 72.0

**Model** AIC 127.61, RMSEA .14

# Grades 6-8



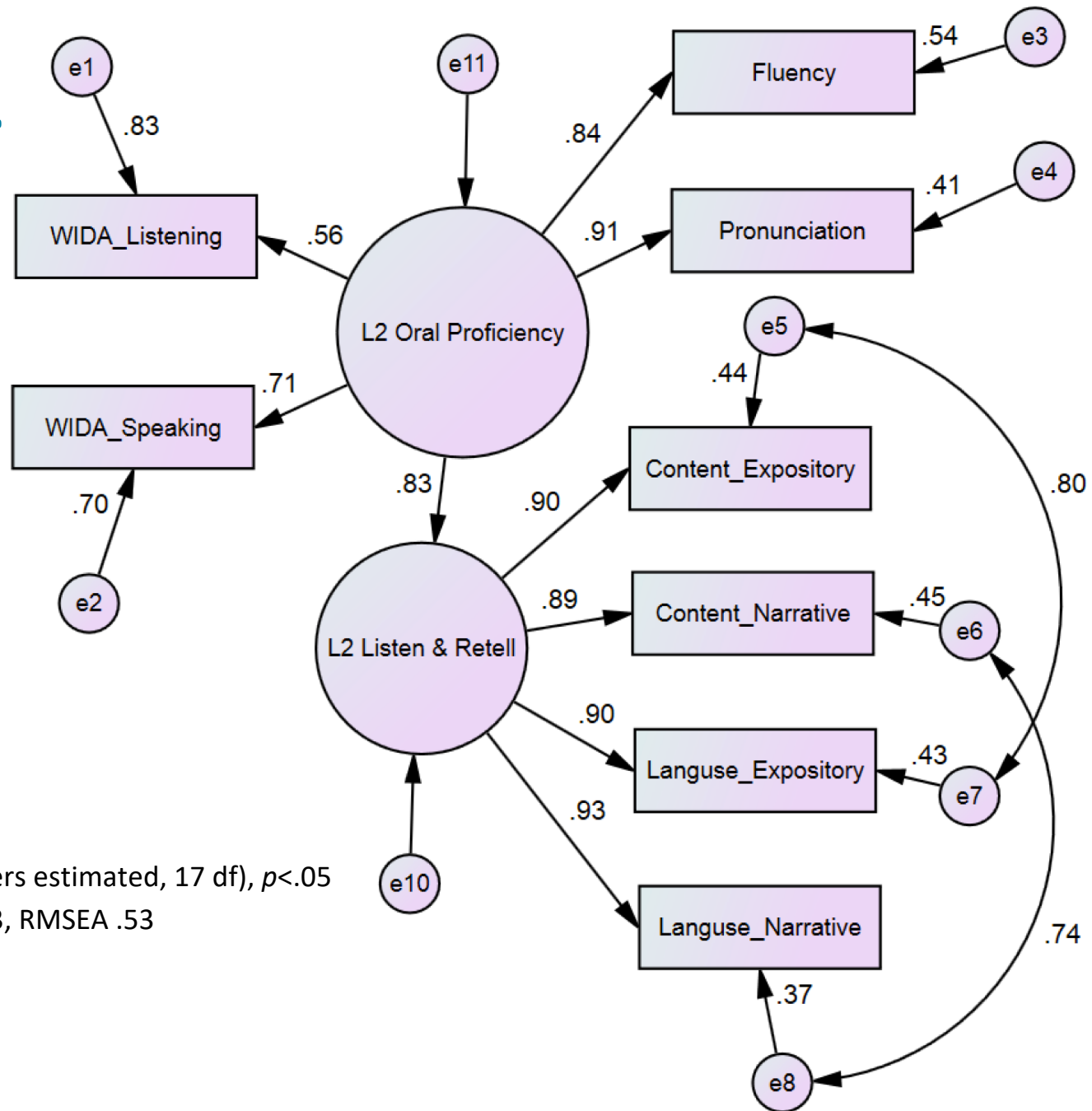
Chi-square = 68.4 (19 parameters estimated, 17 df),  $p < .05$

**Independence model** AIC 1648.51, RMSEA .61

**Saturated model** AIC 72.0

**Model** AIC 106.36, RMSEA .14

# Grades 9-12



Chi-square = 51.3 (19 parameters estimated, 17 df),  $p < .05$

**Independence model** AIC 849.3, RMSEA .53

**Saturated model** AIC 72.0

**Model** AIC 89.29, RMSEA .14

# Relationship of L2 Proficiency measures

- Grades 3-12:
  - Strong relationship between the indicators and the latent variables
  - Strong relationship between the two latent variables
- Grades 1-2
  - External Listening and Speaking << Pronunciation & Fluency
  - Less strong relationship between the two latent variables



## How do ELLs of different grade bands and proficiency levels represent the content of listening passages in their spoken responses?

- Students at higher levels of oral proficiency generally produced more content and received higher scores  
*(The differences among the levels differed by grade bands)*
- Narratives yielded more words than expository prompts and higher scores on content  
*(Across all grade bands besides grades 1-2)*

What is the relationship between L2 oral proficiency and the ability to perform Listen-Retell tasks?

- Strong relationship between L2 oral proficiency and the ability to perform Listen-Retell tasks  
*(Across all grade bands besides grades 1-2)*

# Limitations

- Exploratory
- The prompts presented to test-takers were not controlled
- Language use did not include vocabulary
- Grouped Task 1 (picture) and Task 2 (video) together
- Number of responses for Grades 1-2 was different from other grade bands (Relationship between the two latent variables)

# Implications

- Narrative vs Expository
  - Produce different scores with labels such as “general English” and “academic English”
- Lower grade band
  - Formal assessment
  - Listen and retell task

**ALWAYS LEARNING**

**Saerhim Oh**  
**[sae.oh@pearson.com](mailto:sae.oh@pearson.com)**