

# Tohoku English Language Teaching Expo

## Rater Priorities and the Effectiveness of Rubric Use: An Analysis of Performance Tests for Speaking Interaction in Elementary School Foreign Language Classes

Tsuyoshi SATO, Shunsuke SASAKI, Taiki HATAKEYAMA

# Introduction

- Importance of Communication Skills in English Education
- Challenges in Assessing Speaking Skills
- Difficulties in Objective Speaking Assessment
- Subjectivity in Speaking Tests (Koizumi, 2021)
- Reliability Issues in Speaking Evaluation
- Impact of Non-verbal Factors on Assessment (Sato et al., 2023)
- Overemphasis on Grammar in Speaking Evaluation

# Previous Studies

**Topic:** Challenges and Solutions in Speaking Performance Assessment in Primary English Education

**Focus:** Analysis of speaking assessments in elementary schools and the importance of rubrics in achieving consistent and reliable evaluation.

# Previous Studies

## **Increasing Demand for Speaking Evaluation:**

- Koizumi et al. (2017): Emphasis on balanced 4-skill English education in curriculum guidelines.
- Rise of speaking activities in English classes.
- Problem: Growing demand for appropriate speaking evaluation methods.

# Previous Studies

## **Role of Rubrics in Speaking Evaluation:**

- Use rubrics to maintain evaluation consistency.

## **Benefits of Rubrics:**

- Standardized evaluation across different raters.
- Clearer feedback for students.
- Improved teacher guidance.

# Previous Studies

## Case Study: Kaneko (2022)

- **Problem:** Lack of evaluation of student-to-student interactions in communication activities.
- **Study:** Introduction of Cambridge English Exam methods to assess student interactions.
- **Results:**
  - ✓ Assessment influenced by student pairings.
  - ✓ Low inter-rater reliability among multiple assessors.

# Previous Studies

## Rubric-Based Solutions (Kaneko, 2022)

- **Recommendation:** Create rubrics tailored to the test's purpose and content..
- **Impact:** Establish shared evaluation standards to address inconsistencies.

# Previous Studies

## **Non-verbal Factors in Assessment (Sato et al., 2023)**

**Research Objective:** Analyze the impact of non-verbal factors in speaking assessments.

### **Key Results:**

- Grammar accuracy and eye contact had the strongest influence on evaluations.

**Conclusion:** Over-reliance on grammar accuracy and non-verbal factors may reduce the validity of assessments.

# Previous Studies

## **The Importance of Rubrics (Koizumi, 2021)**

### **Rubric Advantages:**

1. Increased transparency and reliability in scoring.
2. Clear feedback for future learning.
3. Provides insights for teaching improvements.
4. Easier tracking of student progress.
5. Clearer communication of results with stakeholders.

# Aim of this Study

- Investigate how the presence or absence of rubrics impacts speaking performance evaluation in elementary English education.
- Empirical comparison of evaluations with and without rubrics to determine effectiveness.

# Research Questions

**Objective: Compare assessment results in speaking performance tests for elementary foreign language education.**

RQ 1: What differences exist in the criteria emphasized by evaluators when assessing speaking performance with and without a rubric?

RQ 2: How do evaluators approach the assessment process with and without a rubric

# Method

## Participants

- 165 second-year university students from the National University of Education's elementary course.

## Groups

- Rubric Group (r +): 94 participants assessing performance with rubrics.
- No Rubric Group (r -): 71 participants assessing performance without rubrics.

**Background:** All participants have completed coursework in elementary foreign language education and lectures on speaking assessment.

# Method

## Rubric Development

□ Based on Koizumi et al. (2017) and Koizumi (2021).

## □ Key Considerations:

- Assessing the purpose of the test is critical for effective rubric creation.
- Avoid overly detailed penalties for linguistic accuracy to prevent discouraging risk-taking in language use.

# Method

## Rubric Development

- Task Focus: Designed for performance tests at the end of the "Unit 1: Self-Introduction" lesson.
- **Key Components to Evaluate:**
  - Ability to convey personal information to the ALT (task achievement).
  - Knowledge aspects:
    - ✓ **Grammar Accuracy**
    - ✓ **Pronunciation Accuracy**
    - ✓ **Fluency**

# Method

## Rubric Development

- ❑ Can rubric design prevent non-verbal factors (e.g., eye contact) from influencing English evaluations?
- ❑ Excluded Factors:
  - ✓ **Eye contact**
  - ✓ **Gestures**
  - ✓ **Voice volume**

# Method

## Rubric Development

Score	Evaluation Criteria
4	No grammatical mistakes, natural pronunciation that allows the ALT to fully understand the content, and smooth, natural continuation of the conversation.
3	Few grammatical mistakes, mostly natural pronunciation that allows the ALT to broadly understand the content, and generally smooth continuation of the conversation.
2	Some grammatical mistakes, somewhat unnatural pronunciation that makes it difficult for the ALT to fully understand the content, and signs of hesitancy in speech.
1	Many grammatical mistakes, unnatural pronunciation that makes it difficult for the ALT to understand the content, with noticeable silences.

# Method

**6 types of performances emphasizing different aspects:**

- grammar errors (GA) 

- *katakana* pronunciation (PA) 

- pauses (FL) 

- no gestures (GT) 

- no eye-contact (EY) 

- insufficient voice volume (VL) 

# Method

## **6 types of performances emphasizing different aspects:**

- Shared script for the interactions to eliminate the influence of factors other than the six criteria
- Holistic and analytic evaluation
- four-point scale evaluation (4 = Very High, 3 = High, 2 = Low, 1 = Very Low)

# Method

## Data Analysis

- Multiple regression analysis
- Software: JASP (version 0.19.0, JASP Team, 2023)
  - ✓ Dependent variable: holistic evaluation
  - ✓ Independent variables: analytic evaluation for six criteria
- Comparison of analysis results for experiments conducted with and without rubrics.
- Purpose: To examine how rubrics help reduce variability in evaluation across different criteria.

# Method

## Data Analysis

- Decision Tree Analysis
  - ✓ Software: R (version 4.3.1, R Development Core Team, 2023) Purpose:
  - ✓ To further investigate evaluators' decision-making processes.
  - ✓ Comparison of results with the multiple regression analysis.

# Results and Discussion

## Descriptive Statistics

		<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>
<b>GA</b>	<i>r</i> -	3.37	0.99	1.00	4.00
	<i>r</i> +	3.49	1.03	1.00	4.00
<b>PA</b>	<i>r</i> -	3.22	0.92	1.00	4.00
	<i>r</i> +	3.35	0.82	1.00	4.00
<b>FL</b>	<i>r</i> -	3.03	1.03	1.00	4.00
	<i>r</i> +	3.11	1.01	1.00	4.00
<b>GT</b>	<i>r</i> -	3.07	1.07	1.00	4.00
	<i>r</i> +	2.97	1.12	1.00	4.00
<b>EY</b>	<i>r</i> -	3.22	1.07	1.00	4.00
	<i>r</i> +	3.37	1.05	1.00	4.00
<b>VL</b>	<i>r</i> -	3.47	0.68	1.00	4.00
	<i>r</i> +	3.61	0.85	1.00	4.00

# Results and Discussion

## Results of Multiple Regression Analysis: Without Rubric ( $R^2 = .53$ )

Independent Variables	<i>B</i>	<i>SEB</i>	$\beta$
grammar accuracy	0.34	0.02	0.43**
pronunciation accuracy	0.14	0.03	0.16**
fluency	0.17	0.03	0.23**
gesture	0.13	0.02	0.18**
eye-contact	0.28	0.02	0.38**
voice volume	0.13	0.03	0.15**
intercept	- 1.07	0.02	

# Results and Discussion

## Results of Multiple Regression Analysis: With Rubric ( $R^2 = .65$ )

Independent Variables	<i>B</i>	<i>SEB</i>	$\beta$
grammar accuracy	0.51	0.03	0.55**
pronunciation accuracy	0.19	0.04	0.17**
fluency	0.38	0.03	0.41**
gesture	-0.03	0.03	-0.04
eye-contact	0.10	0.03	0.11**
voice volume	0.03	0.03	0.02
intercept	-0.94	0.24	

# Results and Discussion

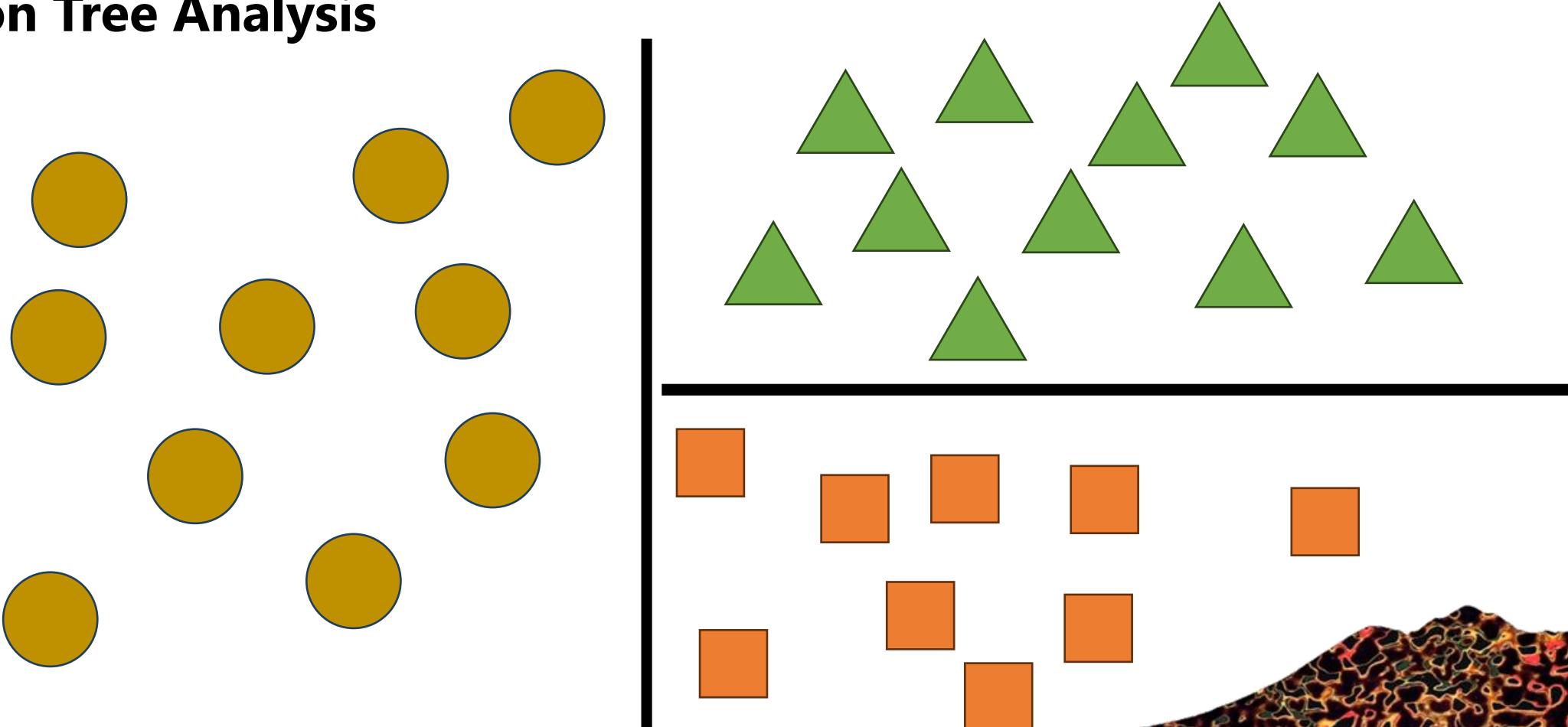
- Excluded Non-Linguistic Factors: Eye Contact, Gestures, Voice Volume → Gestures and Voice Volume were not significant independent variables.
- Indicates the rubric may enhance evaluation validity.
- Eye Contact remained a significant factor, with the standardized beta coefficient decreasing from 0.38 to 0.11, but still influencing the overall evaluation.
- Evaluators still adjusted scores based on eye contact, even though it was not part of the rubric.

# Results and Discussion

- Influence of the 1998 curriculum revision, which emphasized non-verbal communication in foreign language teaching at the elementary level.
- Eye contact has often been a specific focus in communication goals, potentially explaining why it affected scores despite not being included in the rubric.
- Eye contact's impact on overall evaluation should be avoided unless explicitly included in the rubric, to ensure fair and consistent assessments.

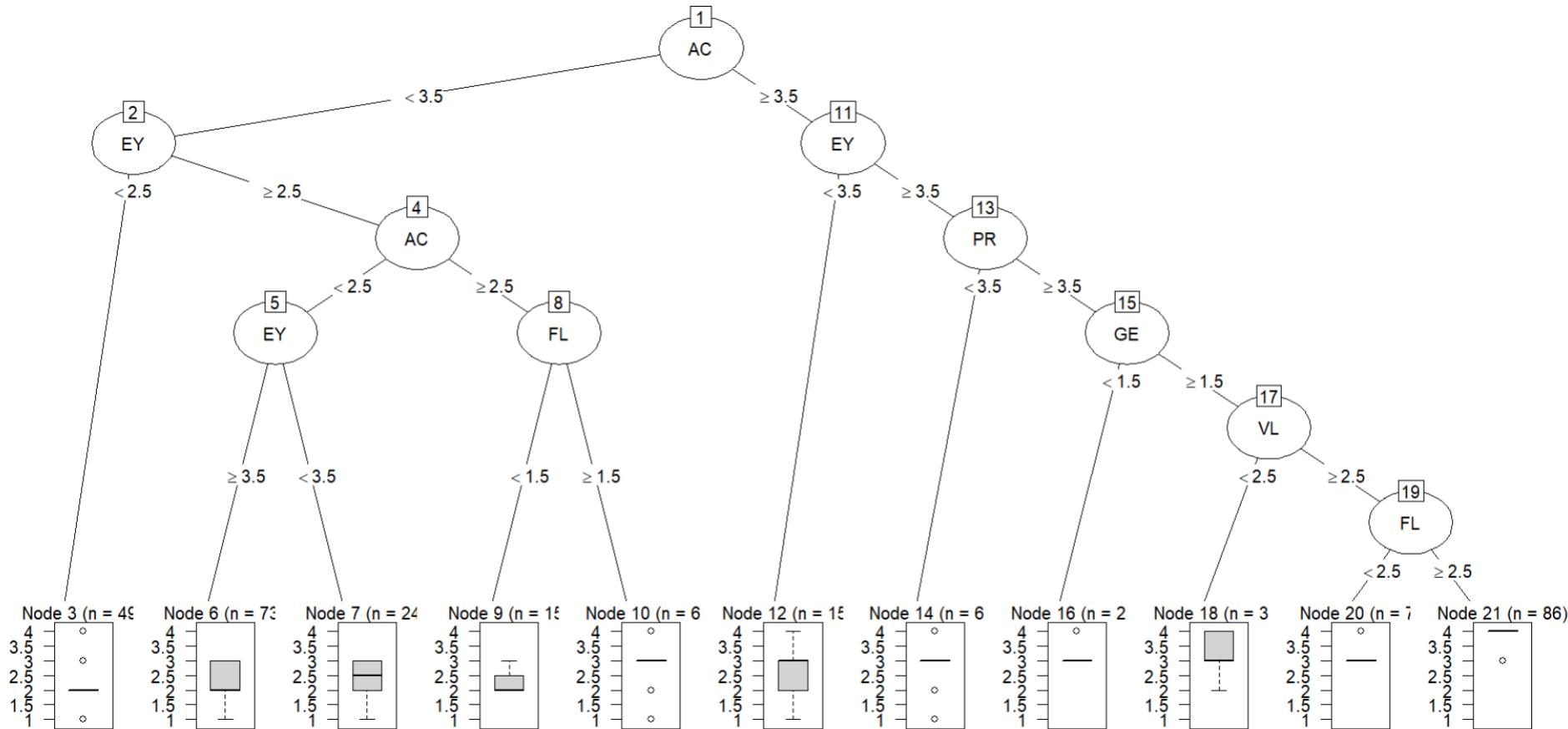
# Results and Discussion

## Decision Tree Analysis



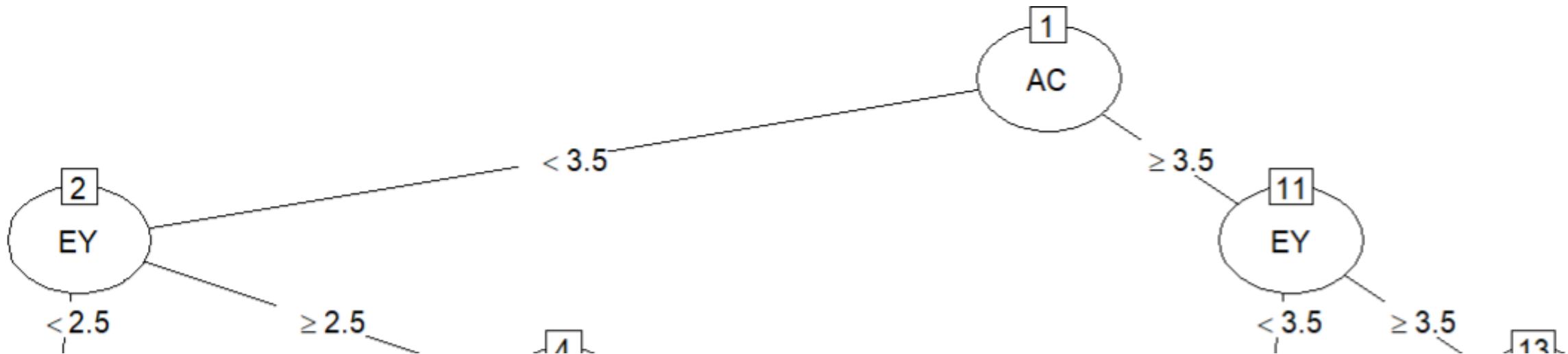
# Results and Discussion

## Decision Tree Analysis: Without Rubric



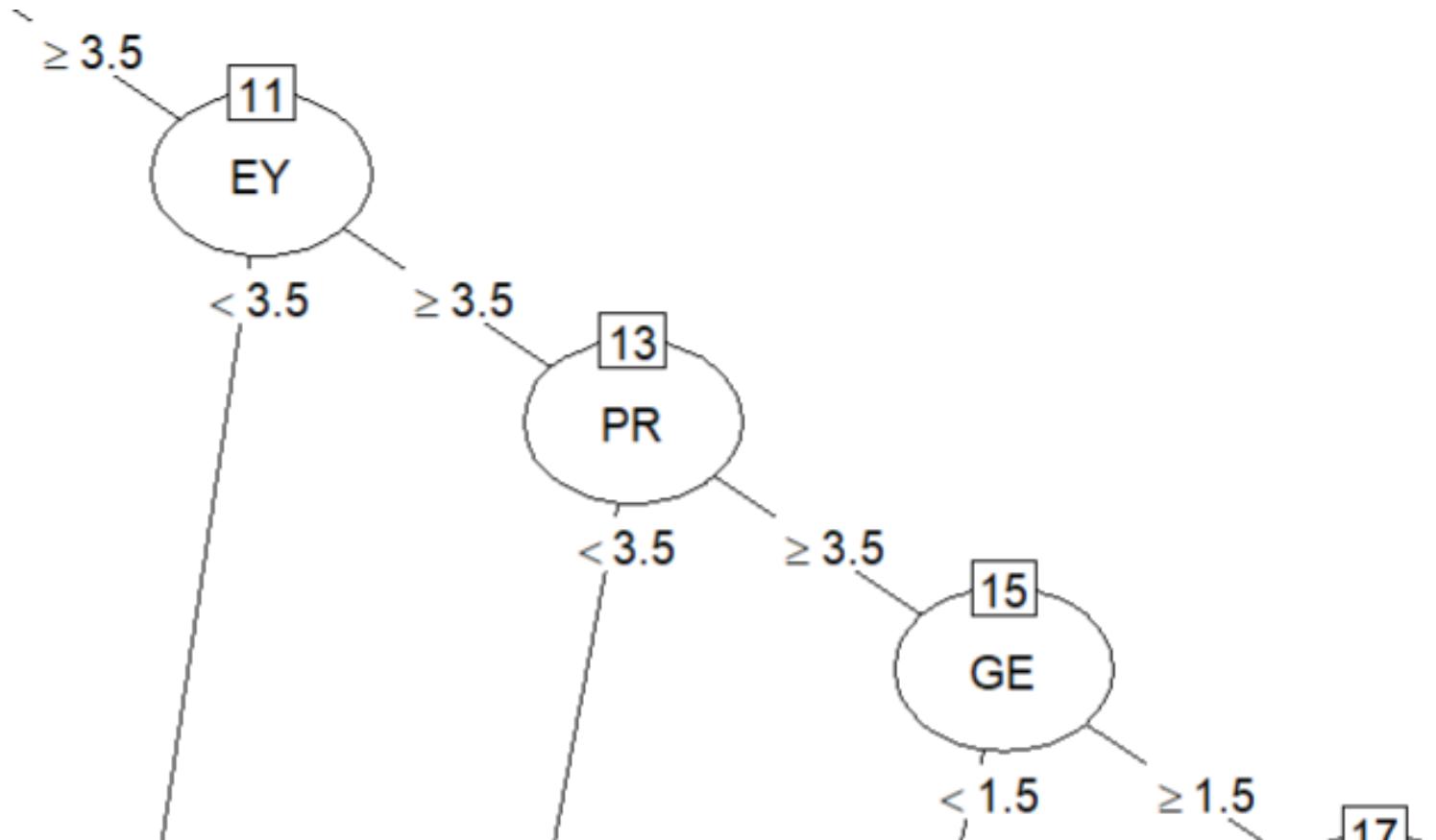
# Results and Discussion

## Decision Tree Analysis: Without Rubric



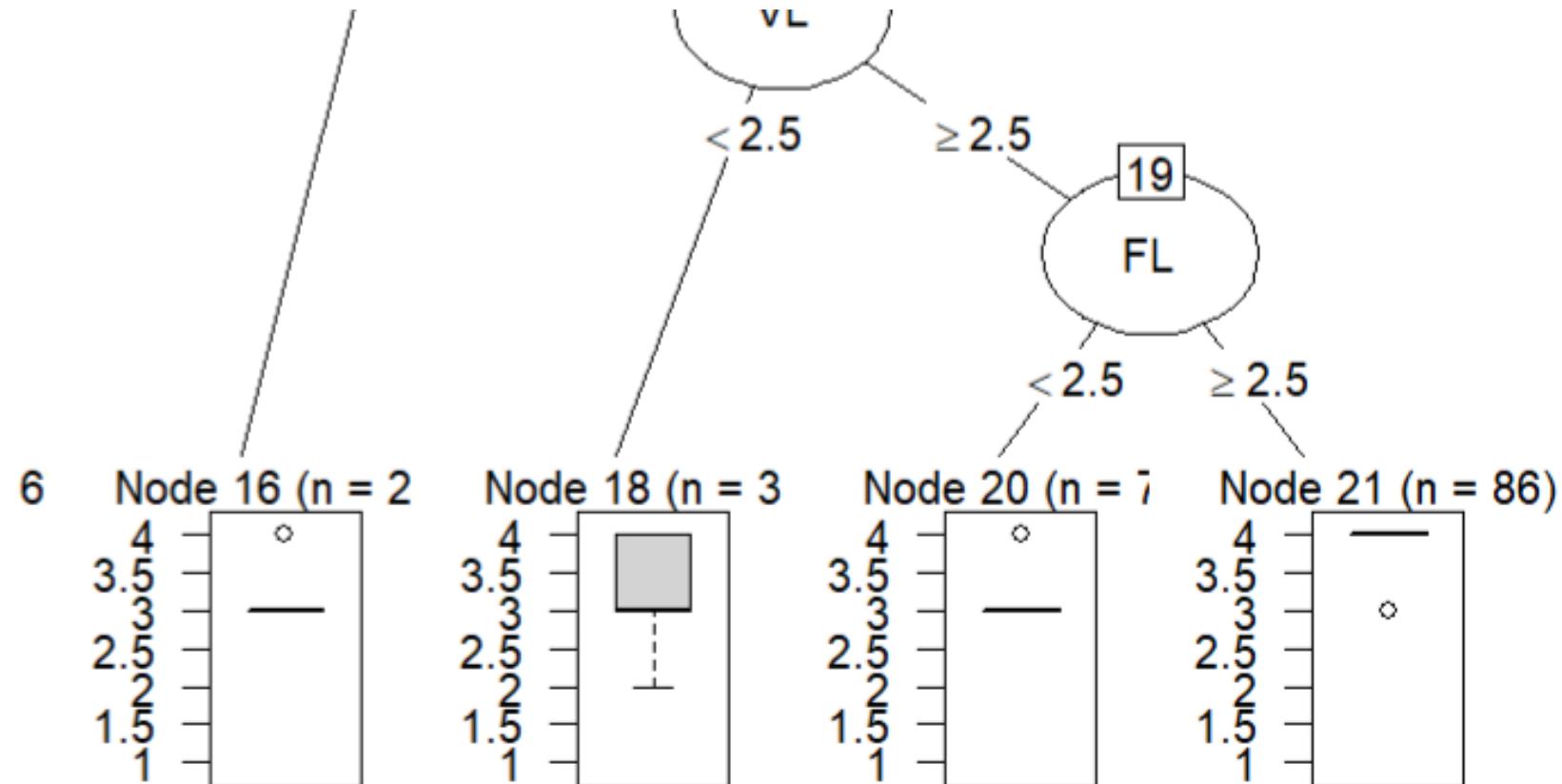
# Results and Discussion

# Decision Tree Analysis: Without Rubric



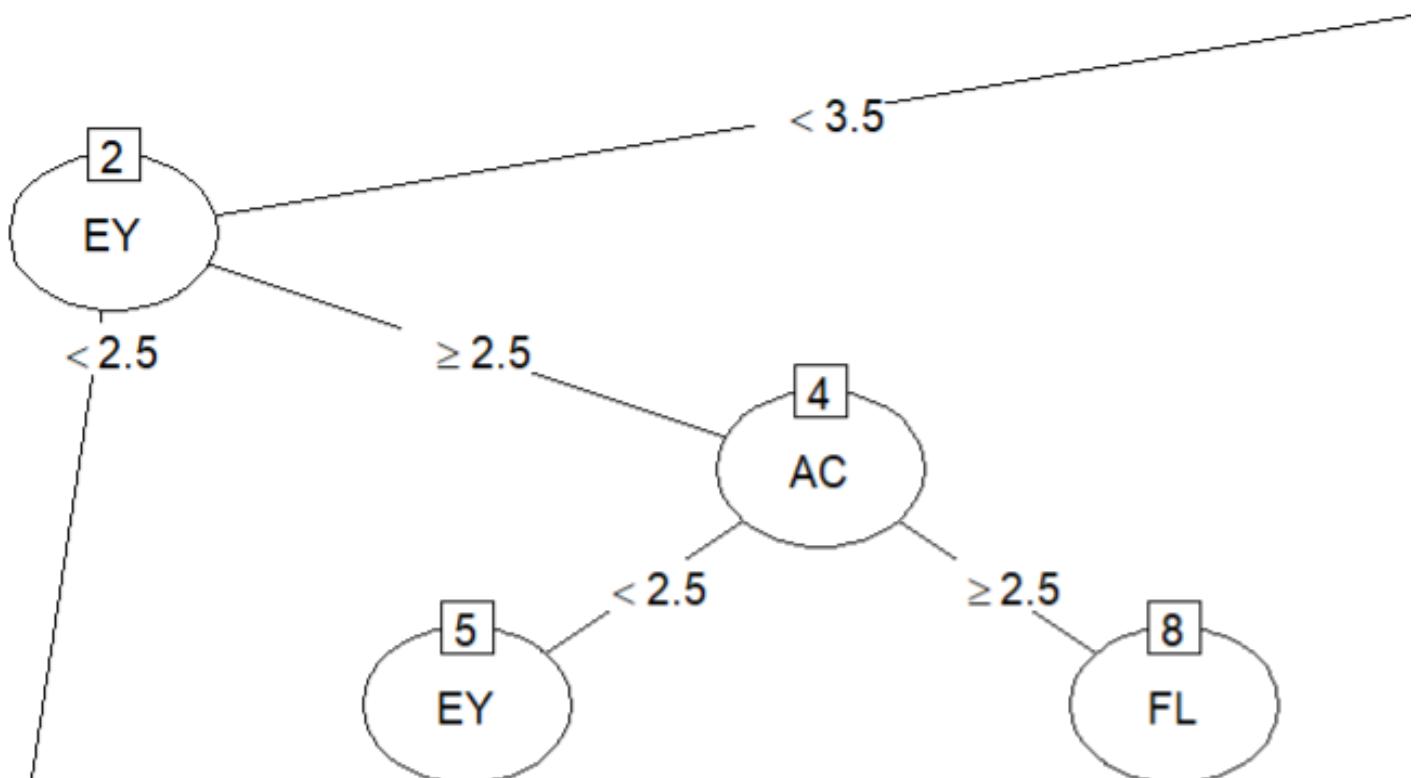
# Results and Discussion

## Decision Tree Analysis: Without Rubric



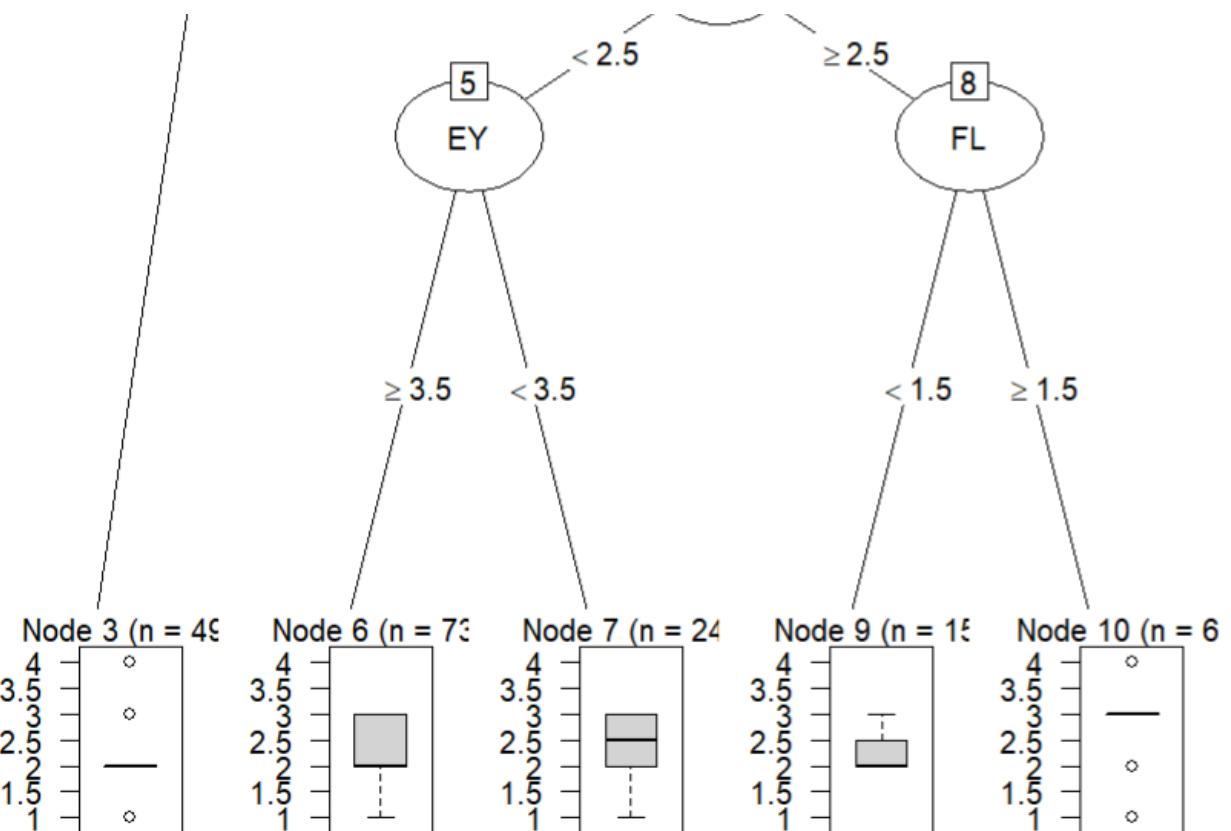
# Results and Discussion

## Decision Tree Analysis: Without Rubric



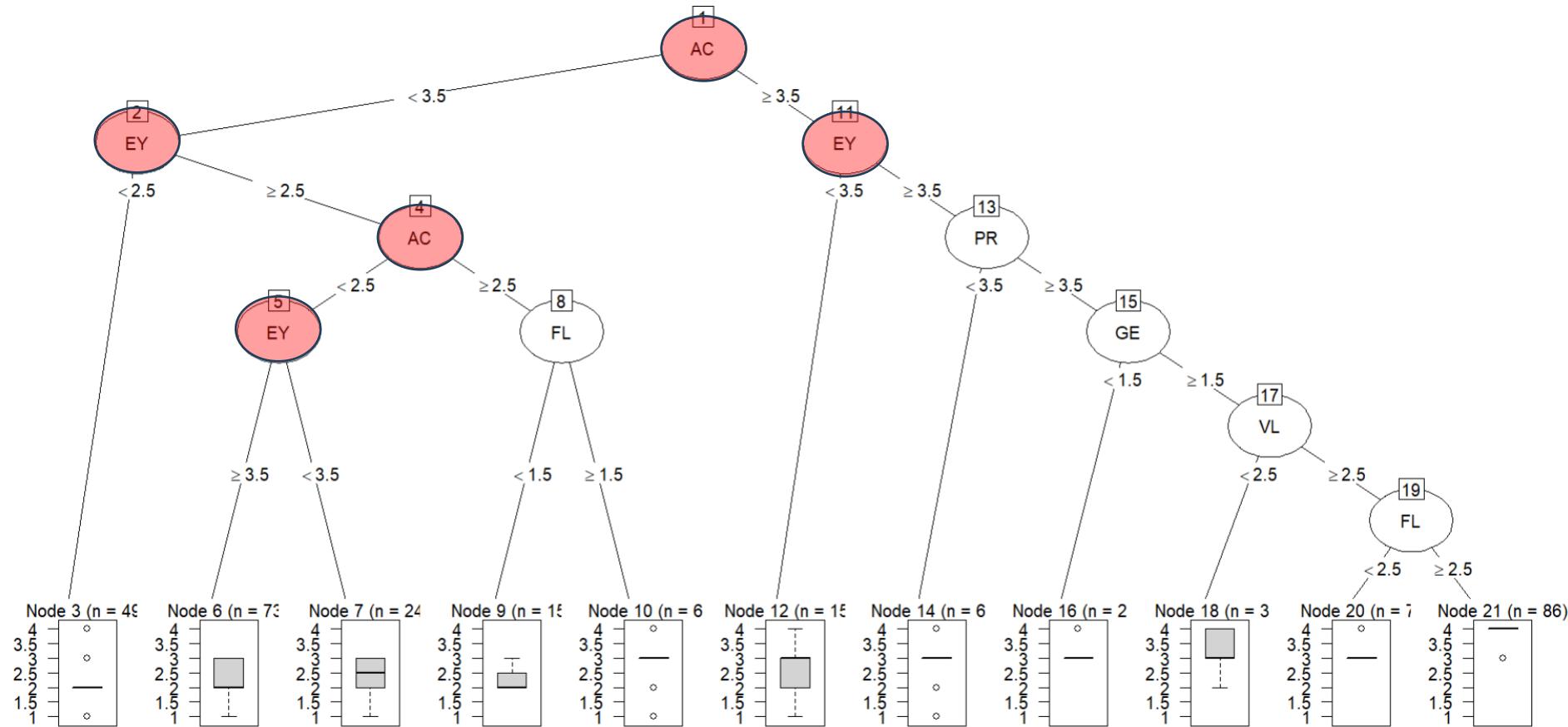
# Results and Discussion

## Decision Tree Analysis: Without Rubric



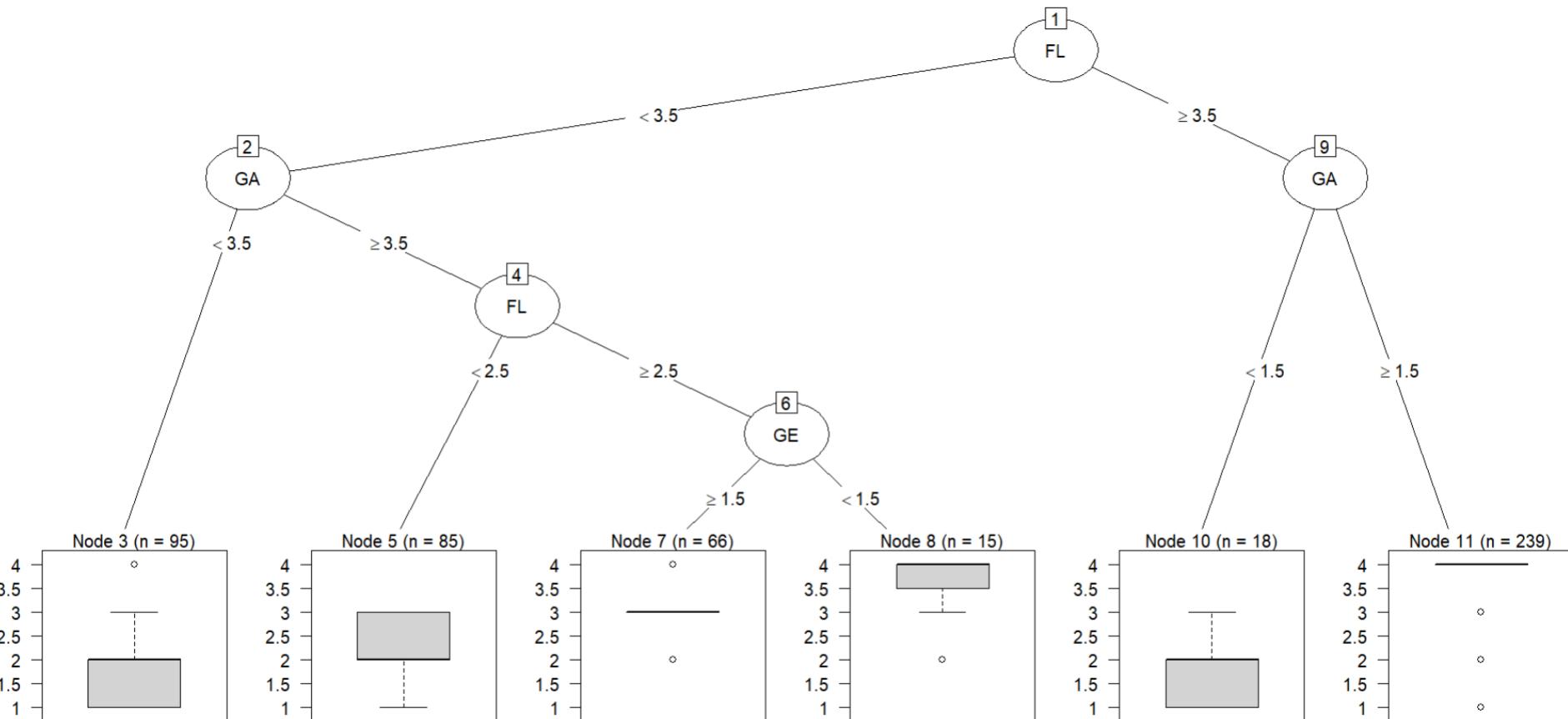
# Results and Discussion

## Decision Tree Analysis: Without Rubric



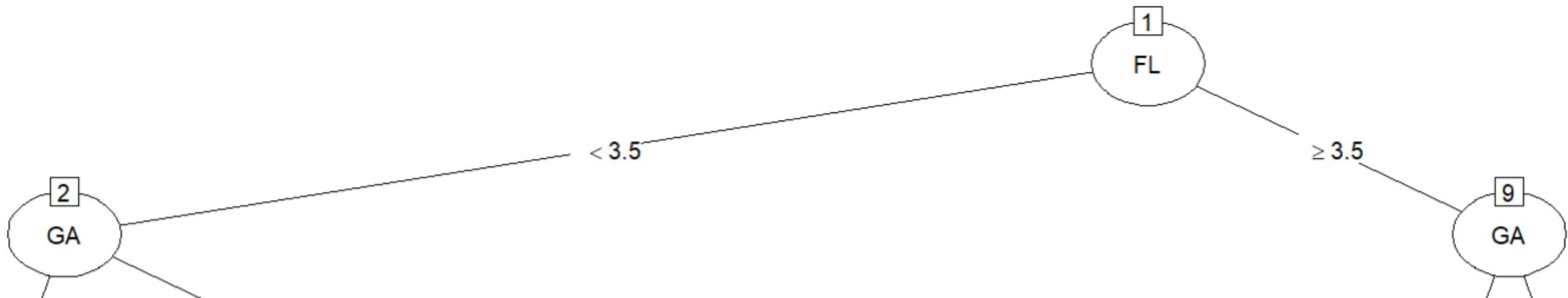
# Results and Discussion

## Decision Tree Analysis: With Rubric



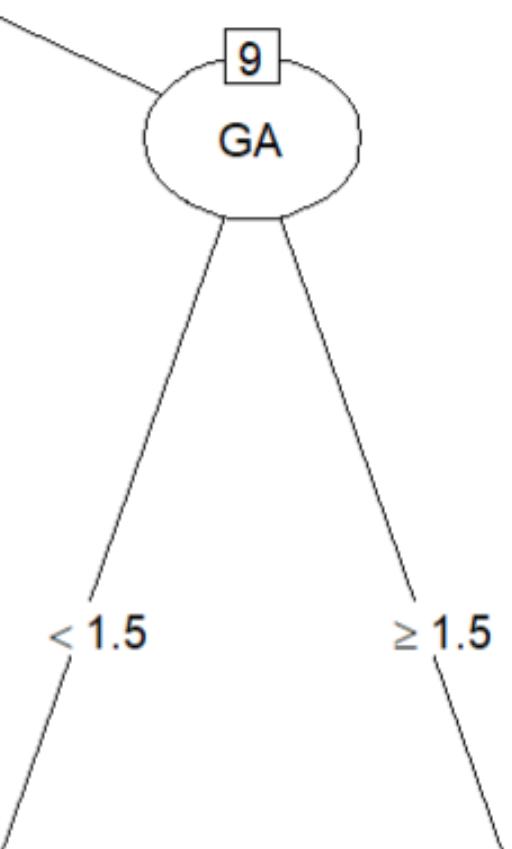
# Results and Discussion

## Decision Tree Analysis: With Rubric



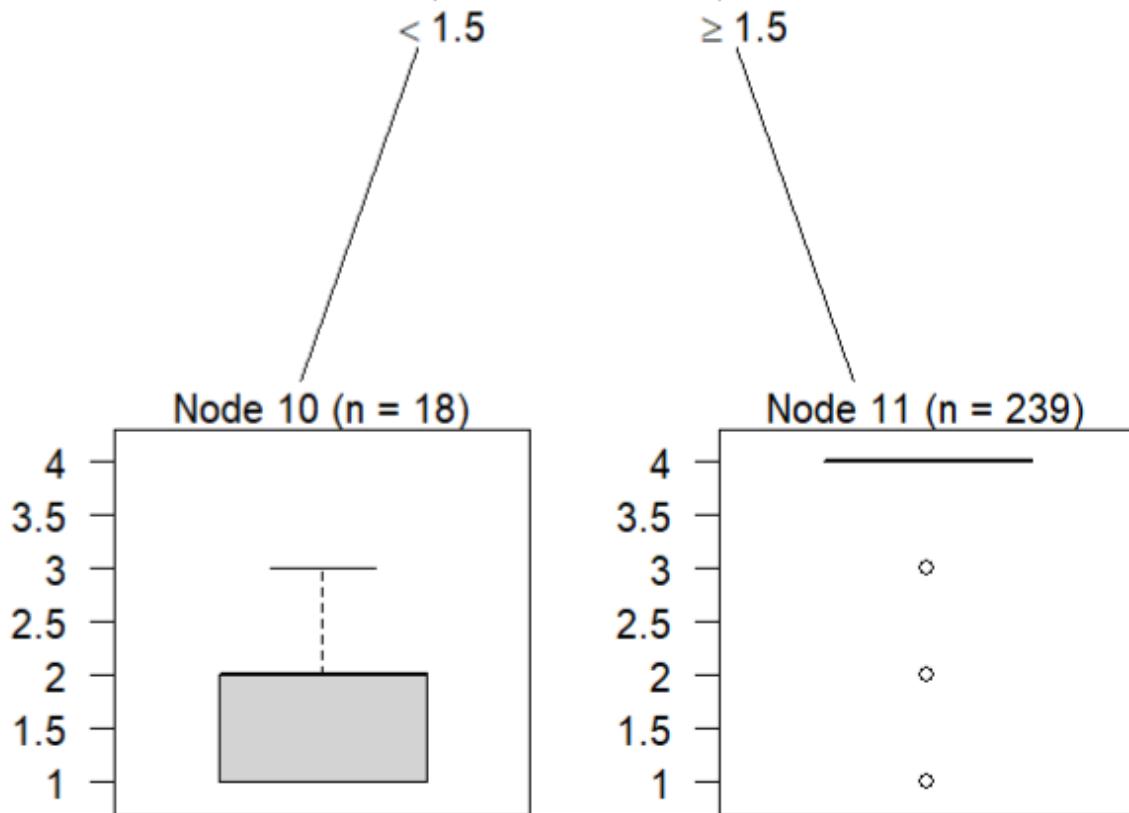
# Results and Discussion

## Decision Tree Analysis: With Rubric



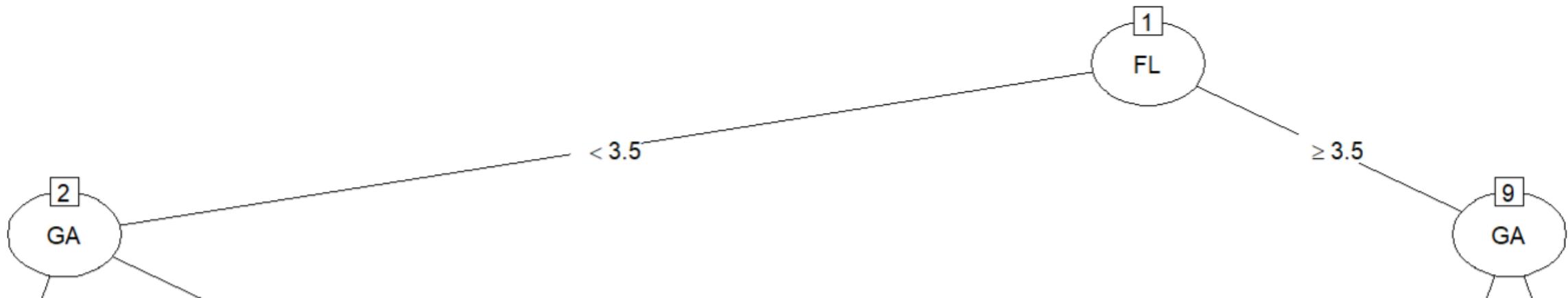
# Results and Discussion

## Decision Tree Analysis: With Rubric



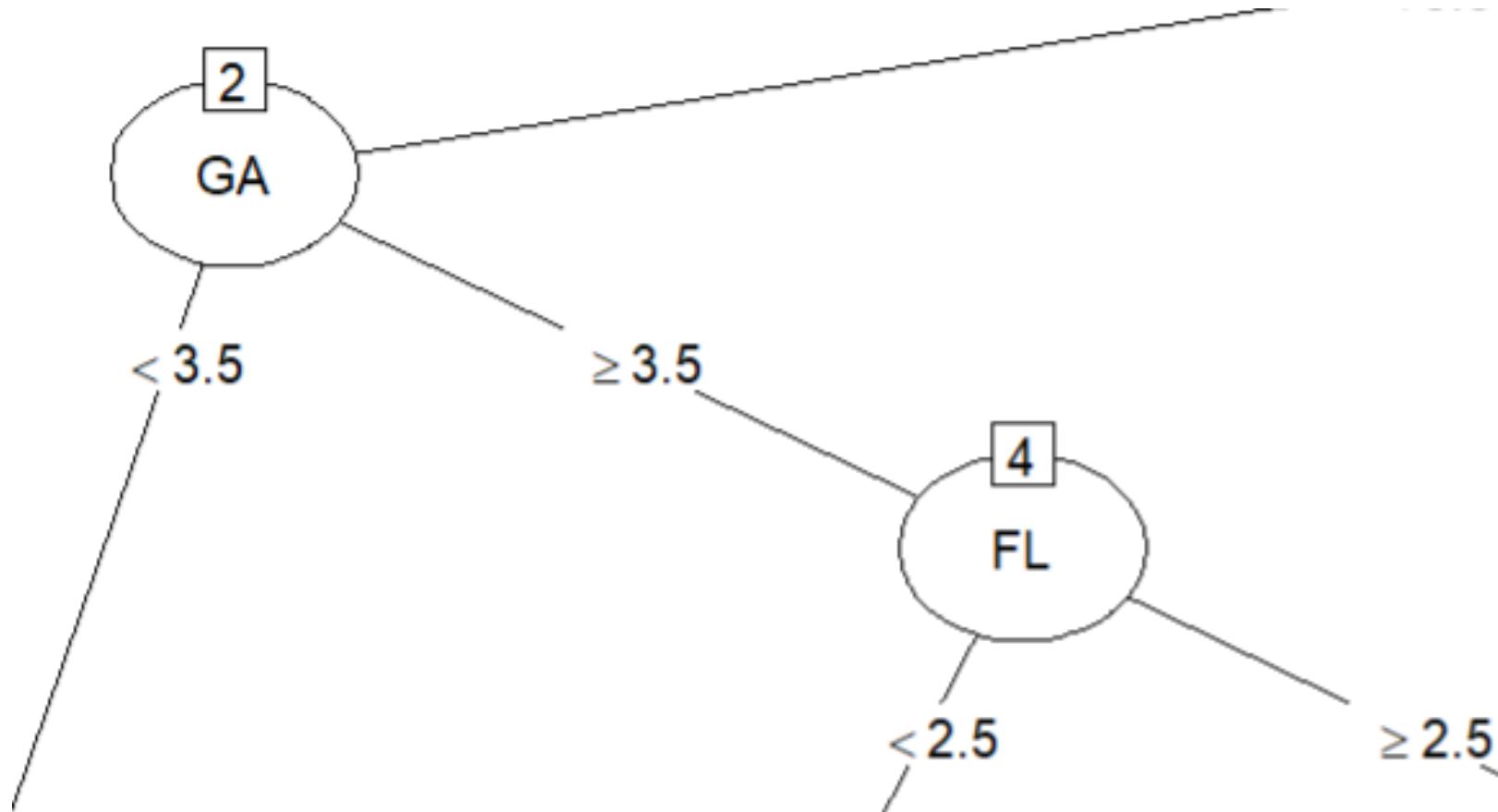
# Results and Discussion

## Decision Tree Analysis: With Rubric



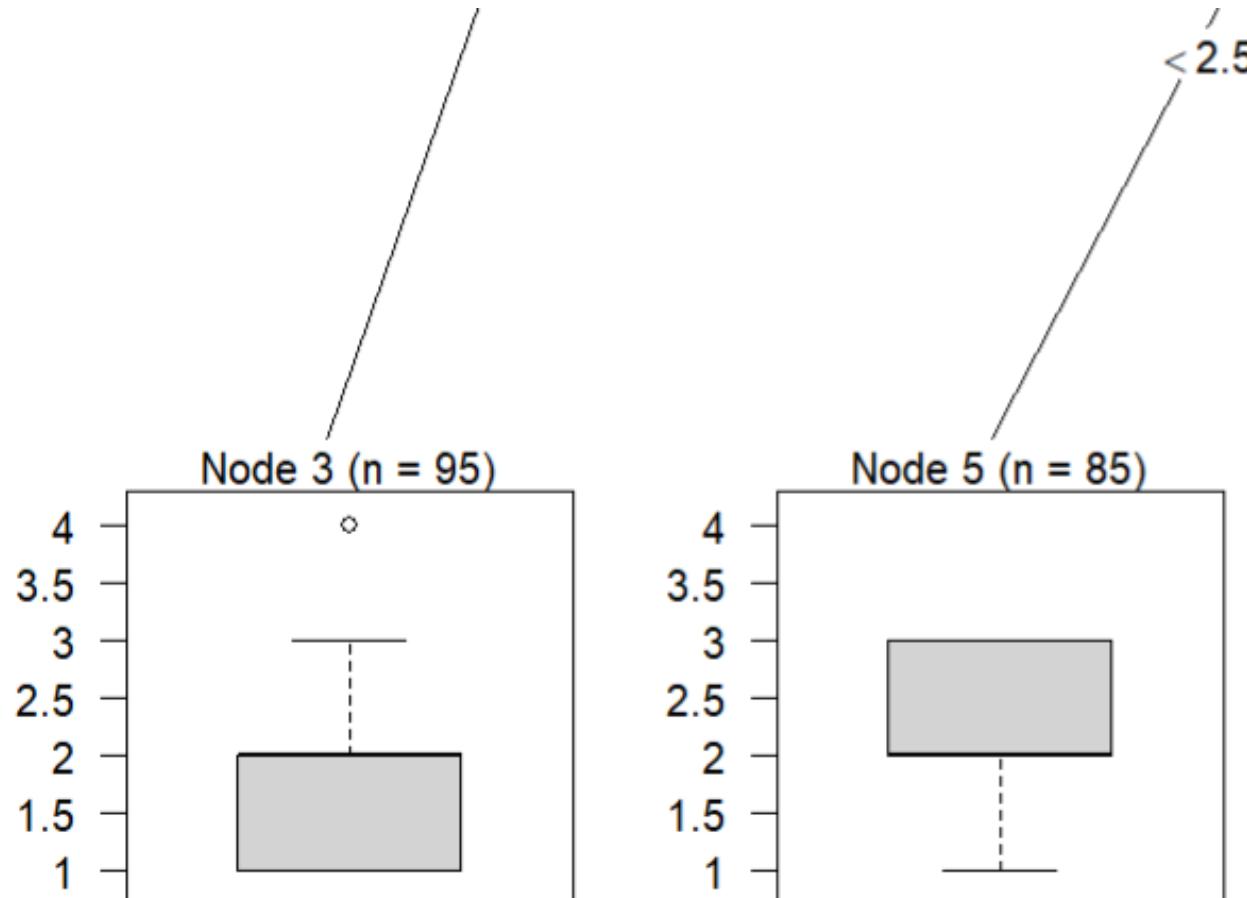
# Results and Discussion

## Decision Tree Analysis: With Rubric



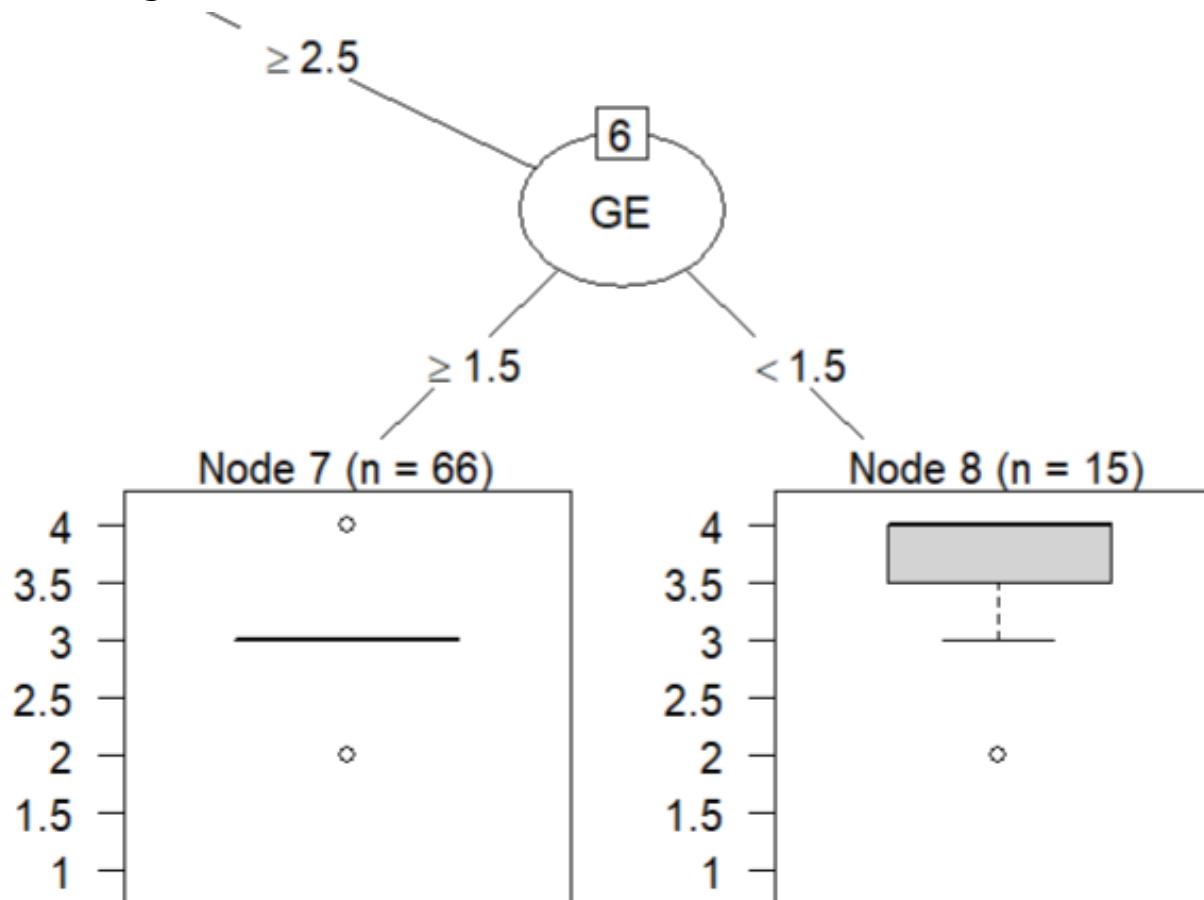
# Results and Discussion

## Decision Tree Analysis: With Rubric



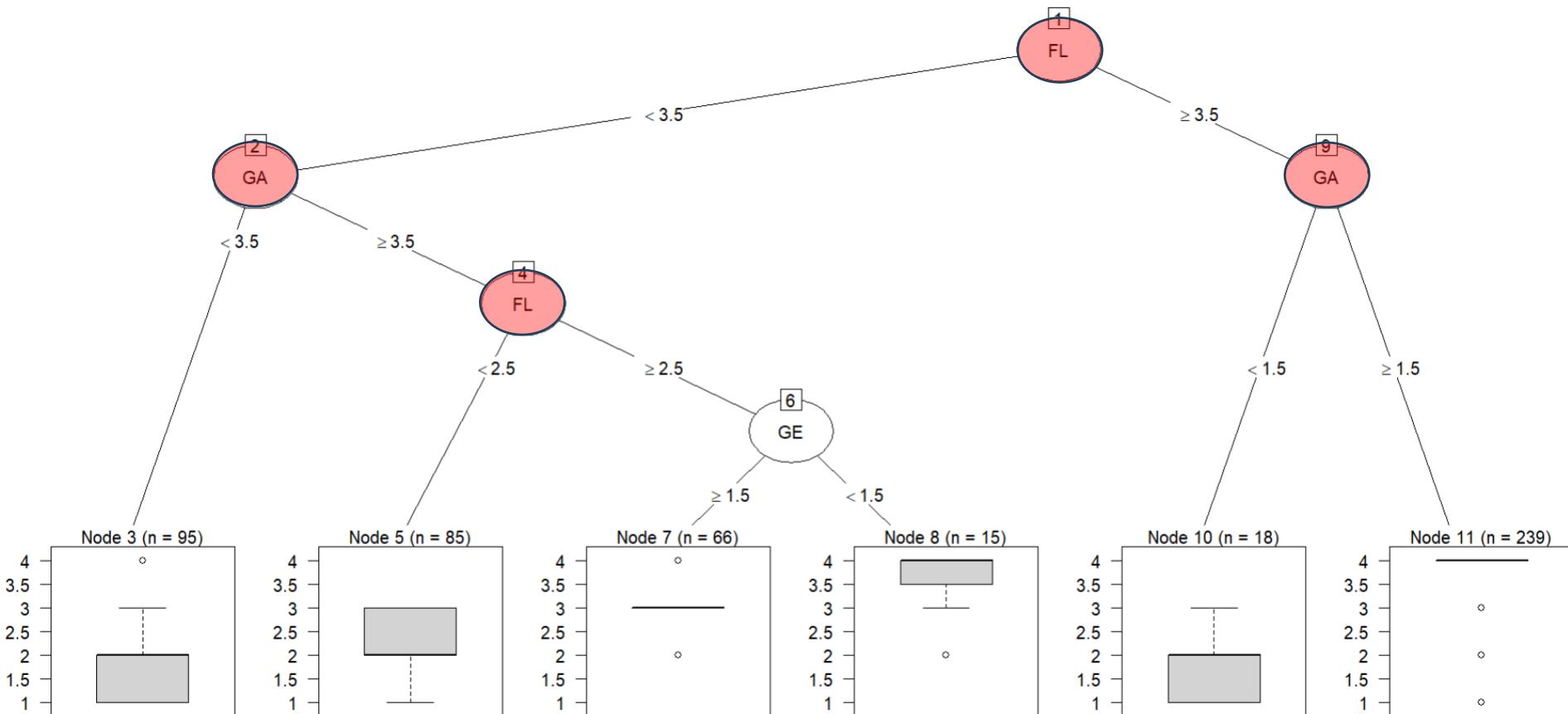
# Results and Discussion

## Decision Tree Analysis: With Rubric



# Results and Discussion

## Decision Tree Analysis: With Rubric



# Conclusion

## Multiple Regression Analysis:

- ❑ Without Rubric:  $R^2 = .53$
- ❑ With Rubric:  $R^2 = .65$
- **Rubric improved the ability to explain overall evaluations based on six factors.**

## Key Influencing Factors:

- ❑ Without Rubric: **grammatical accuracy, eye-contact, fluency, gesture**
- ❑ With Rubric: **grammatical accuracy, fluency, eye-contact**

# Conclusion

## Decision Tree Analysis:

- Without Rubric ⇒ evaluation process becomes complicated; the **same criteria repeatedly referenced**.
- With Rubric ⇒ evaluation process becomes simple and straightforward.
- Help raters **focus on essential evaluation factors**

# Conclusion

## Rubric Effectiveness:

- Focus on Key Criteria: helping distinguish between **critical** and **non-critical** evaluation points.
- Validity and Reliability: improving the fairness of evaluations by reducing the impact of non-specified factors.

# Conclusion

## **Importance of Rubric Development:**

- Clear, specific criteria in the rubric are essential for accurate assessments.
- Collaboration between evaluators is necessary to ensure shared understanding of the rubric.

## **Post-Evaluation Verification:**

- Post-evaluation analysis using methods like multiple regression and decision trees can provide valuable insights for improving rubrics.