

Classification Accuracy of Text Comprehension Tasks for Identifying Children With Vocabulary Delay

Shinyoung Kim^{1,2}, Dongsun Yim^{3*}

¹ Dept. of Communication Disorders, Graduate School, Ewha Womans University, Instructor

² I Say Lab, Speech-Language Pathologist

³ Dept. of Communication Disorders, Graduate School, Ewha Womans University, Professor

Purpose: This study examined text comprehension in children from grades 1 to 6 and investigated the classification accuracy of different text comprehension task conditions for identifying children with vocabulary delay.

Methods: A total of 355 children participated in the study, and 312 children with complete nonverbal IQ data were included in the final analyses. Text comprehension was assessed across presentation modalities (listening, reading-while-listening, and reading), text genres (fiction and expository), and question types (literal and inferential). The seven task conditions derived from these variables were included in the analyses.

Results: Hierarchical logistic regression analyses showed that all seven task conditions significantly contributed to the classification models. In addition, receiver operating characteristic (ROC) and area under the curve (AUC) analyses revealed that the literal condition showed the highest sensitivity, the reading condition demonstrated the highest specificity, and the expository condition provided the most balanced classification performance.

Conclusions: The findings indicated that each task condition served as a meaningful indicator of group differences. Specifically, the literal condition was most useful for screening children at risk (sensitivity=.704), the reading condition was most effective for reducing false classifications of typically developing children (specificity=.916), and the expository condition achieved the most balanced classification performance (AUC=.758, sensitivity=.630, specificity=.888). These findings suggest that task conditions should be selected strategically in text comprehension assessment and intervention according to the purpose of assessment rather than relying on a single task type.

Correspondence: Dongsun Yim, PhD

E-mail: sunyim@ewha.ac.kr

Received: May 31, 2026

Revision revised: July 13, 2026

Accepted: July 31, 2026

This article was based on the first author's doctoral dissertation from Ewha Womans University (2026).

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A3A2A01096102).

ORCID

Shinyoung Kim

<https://orcid.org/0000-0002-2294-2475>

Dongsun Yim

<https://orcid.org/0000-0001-8254-9504>

Keywords: Text comprehension, children with vocabulary delay, discriminative ability, hierarchical logistic regression, ROC curve analysis

1. 서론

텍스트 이해(text comprehension)는 글에 제시된 단어와 문장, 담화 구조로부터 의미를 추출하고, 이를 텍스트를 해석하는 사람의 배경 지식과 통합하여 일관된 정신적 표상으로 구축하는 인지적·언어적 과정이다(Kintsch, 1998; Perfetti, 2007). 다시 말해 텍스트를 이해하기 위해서는 텍스트가 갖고 있는 단어·문장·담화 수준의 정보와 텍스트를 이해하는 주체의 배경 지식 간 상호작용을 통해 텍스트 전체의 의미를 구성할 수 있어야 하며, 이때 사실적 정보에 대한 확인뿐만 아니라 추론과 예측, 요약, 정보의 통합과 같은 상위 인지 기능이 요구된다(Nation, 2019; Oakhill & Cain, 2012).

초등학생의 텍스트 이해 발달을 예측하는 변수로서 다수의 선행 연구들에서는 '어휘 지식'을 보고하고 있는데(Nation, 2005; Ouellette, 2006; Perfetti & Stafura, 2014; Quinn et al., 2015) 여기에는 심층어휘집의 크기와 같은 어휘의 양(breadth)뿐만 아니라 동의어·반의어와 같은 어휘 간 관계, 다의어 처리, 함축적·비유적·관용적 의미 이해, 형태소 인식과 같은 어휘의 질(depth)이 모두 중요한 설명 요인으로 보고되고 있다(Nation, 2005; Ouellette, 2006). 어휘력은 학령기 전반에 걸쳐 텍스트 이해의 핵심적인 설명 요인으로 보고되어 왔으며(Nation, 2005; Ouellette, 2006; Perfetti & Stafura, 2014; Yildirim et al., 2011; Yoon et al., 2018), 어휘력의 공고한 발달 없이는 원활한 텍스트 이해가 어렵다는 점은 다수의 선행 연구들에서 반복적으로 확인되어 왔다.

어휘발달 지연 아동은 텍스트 이해의 핵심적인 언어 능력으로 알려진 어휘 능력에서 포래에 비해 상대적으로 낮은 수행을 보이는 집단이다. 이러한 어휘 능력의 제한은 언어발달 지연(developmental language disorder: DLD) 아동에게서도 흔히 관찰

되는 언어적 특성 가운데 하나이다. 다만 언어발달 지연의 경우 어휘를 포함하여 형태 및 통사, 음운, 담화, 작업기억 등 여러 언어·인지 영역에서의 어려움이 복합적으로 나타나는 이질적인 집단을 포괄하는 개념인 반면(Bishop, 2014; Bishop et al., 2017; Conti-Ramsden & Durkin, 2016; Leonard & Schroeder, 2024), 어휘발달 지연은 언어적 취약성 가운데 단어 의미 지식의 폭과 깊이에 대한 제한이 상대적으로 두드러지는 점이 특징인 것으로 기술되어 왔다(Gray, 2004; McGregor et al., 2013). 이와 같이 언어적 취약성의 특징이 비교적 명확한 집단을 대상으로 할 경우, 언어 전반의 복합적인 요인보다는 특정 언어 능력이 텍스트 이해 수행에 기여하는 방식을 보다 정밀하게 검토할 수 있다는 점에서 이론적·방법론적 측면에서 의미를 갖는다.

언어발달 지연 아동은 전반적인 듣기 이해와 읽기 이해 영역에서 또래에 비해 낮은 수행력을 보이는 경우가 많으며(Bishop & Snowling, 2004; Catts et al., 2003; Stothard et al., 1998), 이러한 어려움에는 어휘 지식의 제한이 중요한 요인으로 작용하는 것으로 알려져 있다. 특히 어휘 능력은 텍스트 내 정보를 통합하거나 명시되지 않은 의미를 추론하는 과정과 밀접하게 관련되어 있어, 어휘 수준의 취약성은 텍스트 이해의 여러 하위 과정에 영향을 미칠 수 있다(Nation & Snowling, 1998; Ouellette, 2006). 따라서 어휘발달 지연 아동을 대상으로 텍스트 이해 특성을 살펴보는 것은 텍스트 이해 과정에서 어휘 지식이 수행하는 역할을 탐색하기 위한 방법론적 접근으로서 타당성을 가진다고 볼 수 있다. 그러나 텍스트 이해를 다루는 국내 연구는 주로 전반적인 언어장애 아동이나 읽기 부진 아동에 초점을 두고 있어, 일반 학급에 재학 중인 어휘발달 지연 아동의 특징적인 텍스트 이해 특성을 독립적으로 분석한 연구는 상대적으로 부족하다. 따라서 어휘발달 지연 아동을 표적 연구 집단으로 설정하여, 이들의 텍스트 이해 능력을 구체적으로 살피고 관련 인지적 요인과의 관계를 검토하는 것은, 학령기 학업 실패 위험군을 조기에 선별하고, 읽기·쓰기 및 교과 학습 지원을 위한 구체적인 중재 목표를 설정하는 데 기초 자료를 제공할 수 있을 것이다.

전통적으로 학령기 아동들을 대상으로 한 텍스트 이해는 주로 ‘읽기’라는 단일 양식을 중심으로 논의되어 왔으나(Cain & Oakhill, 2006; Catts et al., 2003; Gough & Tunmer, 1986; Hoover & Gough, 1990), 실제 학습 환경에서는 청각적 입력과 시각적 입력이 다양한 방식으로 결합되어 제시된다. 교사가 설명을 하면서 읽어주는 교실 상황, 디지털 기반 학습 자료 등은 듣기와 읽기를 함께 활용하는 복합적인 텍스트 환경을 구성하며, 이에 따라 텍스트 이해 역시 단일 양식이 아닌 다양한 입력 조건에서 살펴볼 필요가 있다.

듣기 이해와 읽기 이해는 각각 청각적 입력과 문자 정보를 바탕으로 의미를 구성하는 언어 이해 과정으로 볼 수 있다(Cain & Oakhill, 2008; Gough & Tunmer, 1986; Hoover & Gough, 1990). 한편 듣기와 읽기가 동시에 제공되는 복합 양식의 텍스트 이해는 청각적 정보와 시각적 정보를 통합하여 의미를 구성하는 과정을 의미한다(Verhoeven & Perfetti, 2008). 이러한 복합 양식에서의 텍스트 이해 수행력은 연구에 따라 서로 다른 결과를 보고하고 있다. 듣기+읽기 조건이 단일 양식 조건보다 높은 수행력

을 보였다고 보고한 연구들(Bae, 2025; Chung, Jeong, et al., 2023; Montali & Lewandowski, 1996; Valentini et al., 2018)에서는 공통적으로 복합 양식 조건이 아동의 해독 및 의미 처리 부담을 경감시키고, 청각과 시각의 이중 표상(dual coding)을 통해 정보를 보다 안정적으로 부호화할 수 있도록 한다는 점을 강조한다. 반면, 언어발달 지연 아동과 같이 언어 능력이 취약한 아동에게는 이중과제(dual-task) 또는 복합 양식 처리 상황과 같이 두 가지 정보를 동시에 처리해야 하는 입력 상황이 오히려 인지적 부담으로 작용할 수 있음을 밝힌 선행 연구들도 있다(Hoffman & Gillam, 2004; Leclercq et al., 2013; Yim & Han, 2019). 이러한 상반된 연구 결과들은 복합 양식 입력 상황이 일반 아동에게는 이해를 촉진하는 인지적 이득으로 기능할 수 있지만, 언어 또는 어휘 발달이 지연된 아동에게는 오히려 처리 부담으로 작용할 가능성이 있음을 시사한다.

초등학생들의 텍스트 이해 능력에 영향을 미치는 요인의 하나로, 지문의 종류를 들 수 있다. 초등학교에서 활용되는 텍스트는 크게 인물과 사건을 중심으로 한 문학 지문과 개념이나 사실을 설명하는 비문학 지문으로 구분할 수 있다. 이 두 지문 종류는 내용과 담화 구조, 텍스트를 구성하는 언어적 특성에서 뚜렷한 차이를 보이며, 이는 학령기 아동에게 서로 다른 이해 전략과 인지적 부담을 요구한다(Best et al., 2008; Duke & Roberts, 2010; Graesser et al., 2003). 문학 지문은 서사 구조를 바탕으로 인물의 감정과 의도, 사건 간 인과관계를 이해하는 정서적·추론적 처리를 요구하는 반면(Graesser et al., 1994; Kintsch, 1998; Zwaan & Radvansky, 1998), 비문학 지문은 설명, 비교와 대조, 문제와 해결, 원인과 결과 등 논리적인 구조를 기반으로 개념 간 관계를 파악하고 정보를 조직·통합하는 학문적 언어 능력이 요구된다(Duke & Roberts, 2010; Goldman & Rakestraw, 2000; Meyer & Ray, 2011; Snow, 2010). 이러한 지문 종류 간 차이는 학령기 아동의 텍스트 이해 수행력과 발달 양상에 서로 다른 영향을 미칠 수 있다.

초등학생의 텍스트 이해 능력은 단순한 문자 해독을 넘어 텍스트의 의미를 구성하는 고차원적 과정으로, 크게 사실적 이해(literal comprehension)와 추론적 이해(inferential comprehension)로 구분할 수 있다. 사실적 이해는 텍스트에 명시적으로 제시된 정보를 정확하게 기억하고 재인하는 능력으로, 해독과 기본 어휘 지식, 문장 이해를 바탕으로 비교적 이른 시기부터 발달한다(Cain & Oakhill, 2008; Oakhill et al., 2014). 반면 추론적 이해는 텍스트에 제시되지 않은 정보를 배경지식과 언어적 단서, 맥락을 활용하여 통합하고 해석하는 능력으로, 인과관계 추론과 의미 통합, 상황 모형 구성 등 보다 복잡한 인지·언어적 처리를 요구한다(Cain & Oakhill, 2008; Cain et al., 2004; Kendeou et al., 2009). 선행 연구들에 따르면 학령기 초기에는 사실적 이해가 상대적으로 우세하지만, 학년이 높아짐에 따라 작업기억, 언어 추론 능력, 어휘 깊이와 같은 인지적 자원이 텍스트 이해에 점차 더 크게 관여함으로써 추론적 이해의 중요성이 증가하는 것으로 보고되어 왔다(Biemiller, 2003; Cain & Oakhill, 2008; Ouellette, 2006; Wu et al., 2020). 이러한 결과는 사실적 이해와 추론적 이해가 서로 밀접

하게 관련되어 있으면서도 서로 다른 언어적·인지적 요구를 갖는 텍스트 이해의 하위 영역임을 시사한다.

지금까지의 연구들은 주로 텍스트 양식, 지문의 종류, 질문의 유형 등의 조건에 따른 수행력 차이를 검토하는 데에 초점을 두어 왔으며, 특정 과제 조건이 일반 아동과 언어적 취약성을 가진 아동을 효과적으로 구분할 수 있는지 검토한 연구는 제한적이다. 특히 어휘발달지연 아동의 경우 어휘 지식의 제한으로 인해 텍스트 이해 과정에서 어려움을 경험할 가능성도 높음에도 불구하고, 어떠한 텍스트 조건이 이들의 특성을 가장 민감하게 반영하는지에 대한 실증적 검토는 부족한 실정이다. 따라서 텍스트 이해의 평균 수행력 차이를 확인하는 것을 넘어, 각 과제 조건이 일반 아동과 어휘발달지연 아동을 얼마나 효과적으로 판별할 수 있는지를 살펴보는 것은 임상적 선별과 평가 측면에서 중요한 의미를 갖는다.

실제 초등학생들은 교실에서 다양한 형태의 텍스트를 이해하고 학습해야 한다. 표준화된 어휘 검사는 어휘발달지연 여부를 선별하는 데 유용한 도구이며, 이를 통해 확인된 어휘발달지연 아동들이 구체적으로 어떠한 텍스트 과제 조건에서 더욱 어려움을 경험하는지 파악하는 것은 평가 및 중재 계획을 수립하는 데에 중요한 임상적 정보를 제공할 수 있다. 이에 본 연구는 초등학생들의 텍스트 이해 수행력을 제시 양식, 지문 종류, 질문 유형에 따라 체계적으로 검토하면서, 특히 이러한 구체적인 과제 유형 가운데 일반 아동과 어휘발달지연 아동을 가장 효과적으로 구분하는 과제 조건이 무엇인지 검토하고자 하였다. 이를 위한 구체적인 연구 문제는 다음과 같다.

첫째, 텍스트 제시 양식(듣기, 듣기+읽기, 읽기), 지문 종류(문학, 비문학), 질문 유형(사실적 이해, 추론적 이해)은 일반 아동 집단과 어휘발달지연 아동 집단을 유의하게 판별하는가?

둘째, 텍스트 제시 양식(듣기, 듣기+읽기, 읽기), 지문 종류(문학, 비문학), 질문 유형(사실적 이해, 추론적 이해) 중 일반 아동 집단과 어휘발달지연 아동 집단을 가장 효과적으로 판별하는 텍스트 과제 조건은 무엇인가?

II. 연구 방법

1. 연구 대상

본 연구는 이화여자대학교 생명윤리위원회의 승인을 받은 후 실시되었다(ewha-202406-0007-01). 서울시 소재한 초등학교에 재학 중인 1~6학년 아동 총 355명(남 181명, 여 174명)이 연구에 참여하였으며, 한국판 카우프만 간편지능검사(Korean Kaufman Brief Intelligence Test-second edition: KBIT2, Moon, 2020) 점수 결측 사례 43명을 제외한 312명(남 162명, 여 150명)을 최종 분석 대상으로 하였다. 최종 분석 대상자는 모두 KBIT2 비언어성 지능 표준점수가 85점(-1SD) 이상이었으며, 교사 및 보호자로부터 신경학적 손상, 시·청각 결함, 구강 구조 및 운동 능력 장애가 없다고 보고된 아동이었다. 연구 대상자 중 수용·표현 어휘력 검사(Receptive & Expressive Vocabulary Test:

REVT, Kim et al., 2009) 결과 수용어휘력 또는 표현어휘력이 10%ile 미만인 아동은 어휘발달지연(vocabulary delay: VD) 아동으로, 수용어휘력과 표현어휘력이 모두 10%ile 이상인 아동은 일반(typically developing: TD) 아동으로 선별하였다. 연구 대상 아동의 학년별 분포는 1학년 53명, 2학년 65명, 3학년 61명, 4학년 66명, 5학년 57명, 6학년 53명이었으며, 집단별로는 일반 아동 집단 302명, 어휘발달지연 아동 집단 53명이었다.

연구 대상자의 정보를 학년별, 집단별로 Table 1과 Table 2에 제시하였다. 학년 간 수용어휘력, 표현어휘력, KBIT2 점수에 대한 일원분산분석(one-way ANOVA) 결과, 수용어휘력($F_{(5, 349)}=120.089$, $p<.000$) 및 표현어휘력($F_{(5, 349)}=104.734$, $p<.000$)의 학년 간 차이가 유의하였으며, KBIT2 점수는 학년 간 차이가 통계적으로 유의하지 않았다($p>.05$). Bonferroni 사후 검정 결과, 수용어휘력과 표현어휘력 모두 5학년과 6학년을 제외한 모든 학년 간 차이가 통계적으로 유의하였다($p<.05$). 집단 간 수용어휘력, 표현어휘력, KBIT2 점수에 대한 일원분산분석(one-way ANOVA) 결과, 수용어휘력($F_{(1, 353)}=51.195$, $p<.000$) 및 표현어휘력($F_{(1, 353)}=53.872$, $p<.000$)의 집단 간 차이가 유의하였으며, KBIT2 점수는 집단 간 차이가 유의하지 않았다($p>.05$).

Table 1. Descriptive statistics and one-way ANOVA results for REVT and KBIT2 by grade

Measures	Grade 1 (<i>n</i> =53)	Grade 2 (<i>n</i> =65)	Grade 3 (<i>n</i> =61)	Grade 4 (<i>n</i> =66)	Grade 5 (<i>n</i> =57)	Grade 6 (<i>n</i> =53)	<i>F</i>
REVT-r	80.64 (16.97)	97.34 (17.17)	117.79 (19.18)	129.17 (19.57)	141.47 (19.49)	151.68 (16.74)	120.089***
REVT-e	81.91 (15.91)	93.91 (14.17)	116.08 (20.23)	129.24 (23.04)	144.68 (21.50)	145.21 (19.17)	104.734***
KBIT2	115.08 (12.55)	115.53 (14.81)	119.41 (12.58)	120.43 (9.79)	119.29 (12.99)	120.79 (10.50)	2.004

Note. Values are presented as mean (*SD*). REVT-r and REVT-e are raw scores, and KBIT2 is a standard score.

REVT-r=Receptive & Expressive Vocabulary Test-receptive (Kim et al., 2009); REVT-e=Receptive & Expressive Vocabulary Test-expressive (Kim et al., 2009); KBIT2=Korean Kaufman Brief Intelligence Test-second edition (Moon, 2020).

*** $p<.001$

Table 2. Descriptive statistics and one-way ANOVA results for REVT and KBIT2 by group

Measures	TD (<i>n</i> =302)	VD (<i>n</i> =53)	<i>F</i>
REVT-r	123.93 (28.23)	94.08 (26.78)	51.195***
REVT-e	122.92 (29.21)	92.02 (22.08)	53.872***
KBIT2	119.22 (12.49)	115.44 (11.10)	3.853

Note. Values are presented as mean (*SD*). REVT-r and REVT-e are raw scores, and KBIT2 is a standard score.

REVT-r=Receptive & Expressive Vocabulary Test-receptive (Kim et al., 2009); REVT-e=Receptive & Expressive Vocabulary Test-expressive (Kim et al., 2009); KBIT2=Korean Kaufman Brief Intelligence Test-second edition (Moon, 2020).

*** $p<.001$

2. 검사 도구

1) 텍스트 종류

본 연구에서는 Chung, Song 등(2023)과 Cheong, Jeong 등(2023)의 텍스트 이해 과제를 사용하였다. 해당 과제는 3개의 문학 지문과 3개의 비문학 지문으로 구성되어 있다. 문학 지문은 경기도 소재 고등학교에서 발간한 어린이 대상 창작 동화집의 작품을 바탕으로 연구 목적에 맞게 수정하여 사용하였으며, 비문학 지문은 어린이 신문에 게재된 기사 내용을 토대로 제작하였다. 지문의 선정 및 수정 과정에는 언어병리학과 교수 1인, 박사 과정 연구원 3인, 석사 과정 연구원 2인이 참여하여 내용의 적절성을 검토하였다.

문학 지문은 모두 73개의 문장으로 구성하였으며, 지문 간 난이도 차이를 최소화하기 위하여 평균어절길이를 유사한 수준으로 통제하였다($M=11.02$, $SD=.02$). 비문학 지문은 모두 37개의 문장으로 구성하였으며, 문학 지문과 동일하게 지문 간 평균어절길이가 비슷하게 유지되도록 조정하였다($M=11.50$, $SD=.29$). 이는 지문 종류에 따른 수행 차이가 문장 수준의 언어적 복잡성보다는 장르적 특성을 보다 잘 반영하도록 하기 위함이었다.

2) 텍스트 제시 양식

텍스트는 연구 대상 아동들에게 듣기, 듣기+읽기, 읽기의 세 가지 양식으로 제시되었다. ‘듣기’ 조건에서는 그림과 함께 녹음된 음성 자극을 제시하였으며, 아동은 이를 헤드셋을 통해 청취하였다. ‘듣기+읽기’ 조건에서는 듣기 조건과 동일한 음성 자극과 함께 화면에 텍스트를 제시하여 음성과 텍스트를 동시에 처리하도록 하였다. 제시된 텍스트는 화면당 10줄로 구성하였으며, 아동에게 음성 자극을 들으면서 이를 눈으로 따라 읽도록 지시하였다. ‘읽기’ 조건에서는 음성 자극 없이 텍스트만 제시하였고, 아동이 이를 독립적으로 묵독하도록 하였다. 이때 사용된 텍스트는 듣기+읽기 조건에서 제시된 텍스트와 동일하였다.

3) 텍스트 이해 과제

각 텍스트는 미리 배정된 방식으로 제공되었으며, 연구 대상자에게 텍스트가 제시된 직후 해당 지문에 대한 이해 과제를 실시하였다. 텍스트 이해 문항은 언어병리학과 교수 1인, 박사 과정 연구원 3인, 석사 과정 연구원 2인이 참여하여 개발하였으며, 이후 1급 언어재활사 2명을 대상으로 문항의 내용 타당도를 평가하였다. 평가는 5점 척도를 사용하였으며, 모든 문항이 4점 이상의 평가를 받아 적절한 수준의 내용 타당도를 확보한 것으로 판단되었다.

텍스트 이해 과제는 사실적 이해와 추론적 이해를 평가하는 문항으로 구성하였다. 사실적 이해 문항은 총 10개 문항, 추론적 이해는 총 8개 문항으로서, 두 가지 이해 과제 모두 총점은 16점으로 동일하도록 설정하였다. 사실적 이해 문항은 요구되는 정보 처리 수준에 따라 1점과 2점 문항으로 구분하였으며, 1점 문항은 텍스트에 직접 제시된 정보를 확인하는 수준의 문항, 2점 문항은 지문 내 제시된 정보들을 종합하거나 문단 수준의 의미를 파악하여 응답해야 하는 문항으로 구성하였다. 추론적 이해 문항은 텍스트에 명시되지 않은 내용을 지문 내 단서와 배경지식을 활용하

여 추론하도록 설계하였다. 채점은 문항별 배점 기준에 따라 실시하였으며, 이후 정반응률을 산출하여 통계 분석에 활용하였다. 지문에 포함된 사실적 이해 및 추론적 이해 문항의 예는 Appendix 1에 제시하였다.

3. 연구 절차

연구 대상 아동에 대한 모든 검사는 아동이 재학 중인 학교에서 실시하였다. 검사는 외부 자극이 최소화된 독립된 공간에서 연구자와 아동이 1:1로 착석하여 진행하였다. 선별 검사 및 텍스트 이해 과제 모두 동일한 환경에서 실시하였으며, 과제 수행에 필요한 자극은 모두 컴퓨터 모니터를 통해 제시하였다.

텍스트 이해 과제에서는 각 아동이 총 3개의 지문을 듣거나 읽게 되며, 지문은 ‘듣기’, ‘듣기+읽기’, ‘읽기’ 세 가지 제시 양식에 각각 하나씩 배정되었다. 6개의 전체 지문 중 문학 지문과 비문학 지문이 2:1 또는 1:2가 되도록 구성한 후, 각 아동에게 세 가지 제시 양식이 무선으로 할당되었다. 제시 양식별로 사전에 정해진 절차에 따라 텍스트가 제시되었으며, 그 직후 해당 지문에 대한 이해 과제를 실시하였다. 이러한 절차를 통해 시간이 경과함에 따라 발생할 수 있는 기억 소실이나 외부 정보의 영향을 최소화한 상태에서 연구 대상 아동들의 텍스트 이해 능력을 측정하고자 하였다. 이해 문항은 PPT 파일로 제작하여 모니터에 제시하였으며, 문항에 대한 읽기 부담이 수행에 영향을 미치지 않도록 연구자가 모든 문항을 읽어주었다. 아동은 구어로 답하였으며, 연구자는 응답 내용을 즉시 기록하였다. 또한 채점의 정확성과 신뢰도 검토를 위하여 모든 반응을 녹음하여 보관하였다. 텍스트 이해 과제의 전체 소요 시간은 아동의 수행 수준에 따라 차이가 있었으며, 평균적으로 약 40~50분이 소요되었다.

선별 검사와 텍스트 이해 과제는 동일한 회기에서 순차적으로 실시하였으며, 전체 검사 시간은 약 1시간에서 1시간 30분이었다.

4. 신뢰도

텍스트 이해 과제의 채점 신뢰도를 검증하기 위하여 전체 자료 중 10%를 무작위로 선정하여 언어병리학과 박사 과정 연구원 3인이 독립적으로 재검을 실시하였다. 평가자 간 일치도를 분석한 결과 신뢰도 계수는 $r=.90$ 으로 나타나 높은 수준의 일치도를 보였다(DeVellis, 2017; Nunnally & Bernstein, 1994).

5. 결과 처리

본 연구에서는 텍스트 제시 양식, 지문 종류, 질문 유형에 따른 집단 판별 변인을 검토하기 위하여 SPSS 25.0을 사용하여 위계적 로지스틱 회귀분석(hierarchical logistic regression)을 실시하였으며, 이후 R 4.5.1을 사용하여 ROC(receiver operating characteristic) 곡선 분석 및 AUC(area under the curve)를 산출하였다. 또한 일반 아동 집단과 어휘발달 지연 아동 집단 간 표본 수의 불균형을 고려하여, 로지스틱 회귀분석 결과와 함께 분류

정확도, ROC 곡선 분석, AUC, 민감도 및 특이도를 제시하여 각 모형의 판별 성능을 평가하였다.

III. 연구 결과

본 연구에서는 텍스트 제시 양식(듣기, 듣기+읽기, 읽기), 지문 종류(문학 지문, 비문학 지문), 질문 유형(사실적 이해, 추론적 이해)에 따른 집단 판별의 진단적 유용성을 검토하기 위하여 각 변수별 수행력을 독립적으로 투입하여 위계적 로지스틱 회귀 분석(hierarchical logistic regression)을 실시하였다. 이때 모든 분석에서 연령과 KBIT2 점수는 공변량으로 통제하여 분석하였다. 연령과 KBIT2 점수를 투입한 1단계 모형은 통계적으로 유의하였으며($\chi^2=6.473$, $p=.039$), Hosmer-Lemeshow 검정 결과 모형의 적합성을 확인하였다($p>.05$). 이후 각 분석에서는 2단계에서 예측 변인을 개별적으로 투입하여 모형 변화를 검토하였다.

로지스틱 회귀 분석은 결측값 처리 방식으로 리스트 삭제(listwise deletion)를 적용하므로, 연구 대상 아동들 가운데 KBIT2 점수에서 결측치가 존재한 43명의 사례는 분석에서 제외되었으며, 총 312명(TD=285명, VD=27명)이 최종 분석에 포함되었다. 또한 로지스틱 회귀모형의 영향 사례를 확인하기 위하여 Cook's distance를 검토한 결과, 모든 모형에서 Cook's distance가 1 미만으로 나타나 분석 결과에 중대한 영향을 미치는 사례는 확인되지 않았다.

1. '듣기' 조건 수행력의 판별 정확도

텍스트 제시 양식 중 '듣기' 조건 수행력의 예측 요인을 확인하기 위해 2단계에서 '듣기' 조건 수행력 변수를 추가한 결과, 전체 모형은 통계적으로 유의하였으며($\chi^2=12.766$, $p=.005$), -2LL값은 177.267에서 170.974로 감소하여 모형 적합도가 향상되었다. 특히 2단계에서 설명력이 Nagelkerke R^2 기준 .046에서 .090으로 증가하였으며, 따라서 투입 변수인 '듣기' 조건 수행력이 모델 적합도를 유의하게 향상시키는 것으로 확인되었다($\Delta\chi^2=6.293$, $p=.012$). Hosmer-Lemeshow 검정 결과 역시 모형의 적합도가 양호한 것으로 나타났다($p>.05$). '듣기' 조건 수행력을 투입한 위계적 로지스틱 회귀모형의 적합도 분석 결과는 Table 3에 제시하였다.

Table 3. Model fit statistics for the hierarchical logistic regression model of listening condition

Step	χ^2	df	p	-2LL	Cox & Snell R^2	Nagelkerke R^2	Hosmer-Lemeshow test (p)
Step 1	6.473*	2	.039	177.267	.021	.046	.144
Step 2	12.766**	3	.005	170.974	.040	.090	.595
$\Delta\chi^2$	6.293*	1	.012				

* $p<.05$, ** $p<.01$

개별 변수의 회귀 계수를 분석한 결과, 연령의 회귀 계수가 유의하였으며($B=.025$, Wald=5.208, $p=.022$), 승산비는 1.026으로 나타났다. 로지스틱 회귀분석에서 승산비(odds ratio)는 독립변수가 1단위 증가할 때 사건이 발생할 승산(odds)이 얼마나 변화하는지를 나타내는 지표로서, 회귀계수 B 를 지수 변환한 값인 $Exp(B)$ 로 정의된다. $Exp(B)$ 가 1보다 큰 경우에는 독립변수가 1단위 증가할 때 사건 발생 승산이 증가하며, $Exp(B)$ 가 1보다 작으면 사건 발생 승산이 감소하는 것을 의미한다(Hosmer et al., 2013). 본 연구의 결과, '듣기' 조건 수행력을 투입한 위계적 로지스틱 회귀분석에서 승산비가 1.026으로 나타난 것은, 연령이 1단위(개월) 증가할 때 대상자가 VD로 분류될 승산이 약 2.6% 증가함을 의미한다.

또한 '듣기' 조건 수행력의 회귀 계수가 유의하였으며($B=-.027$, Wald=6.384, $p=.012$), 승산비는 .974로, '듣기' 조건 수행력이 1단위 증가할 때 대상자가 VD로 분류될 승산이 약 2.6% 감소하는 것으로 나타났다. 한편 KBIT2 점수는 2단계에서 통계적으로 유의하지 않아($p>.05$), 본 모형에서 집단 판별에 유의한 기여를 하지 않는 것으로 나타났다. '듣기' 조건 수행력에 대한 로지스틱 회귀분석 결과는 Table 4에 제시하였다.

Table 4. Hierarchical logistic regression results for listening condition

Predictor	B	SE	Wald	df	p	$Exp(B)$
Age	.025*	0.011	5.208	1	.022	1.026
KBIT2	-.029	0.017	3.022	1	.082	0.971
Listening accuracy	-.027*	0.011	6.384	1	.012	0.974
Constant	.052	2.083	0.001	1	.980	1.053

Note. KBIT2=Korean Kaufman Brief Intelligence Test-second edition (Moon, 2020).

* $p<.05$

로지스틱 회귀 모형의 분류 정확도를 확인한 결과, 전체 분류 정확도는 91.7%로 나타났으며, 집단별로 살펴보면 TD 집단은 100%가 정확하게 분류된 반면, VD 집단은 3.7%만이 정확하게 분류되었다(Table 5).

Table 5. Classification table for the logistic regression model of listening condition

Observed	Predicted		Percentage correct (%)
	TD	VD	
TD	285	0	100.0
VD	26	1	3.7
Overall percentage	-	-	91.7

Note. TD=typically developing children; VD=children with vocabulary delay.

2. ‘듣기+읽기’ 조건 수행력의 판별 정확도

텍스트 제시 양식 중 ‘듣기+읽기’ 조건 수행력의 예측 요인을 확인하기 위해 2단계에서 ‘듣기+읽기’ 조건 수행력을 변수로 추가한 결과, 전체 모형은 통계적으로 유의하였고($\chi^2=28.075$, $p=.000$), -2LL값은 177.267에서 155.665로 감소하여 모형 적합도가 향상되었다. 2단계에서 설명력이 Nagelkerke R^2 기준 .046에서 .193으로 증가하였으며, 따라서 투입 변수인 ‘듣기+읽기’ 조건 수행력이 모델 적합도를 유의하게 향상시키는 것으로 확인되었다($\Delta\chi^2=21.602$, $p=.000$). Hosmer-Lemeshow 검정 결과 역시 모형의 적합도가 양호한 것으로 나타났다($p>.05$). ‘듣기+읽기’ 조건 수행력을 투입한 위계적 로지스틱 회귀모형의 적합도 분석 결과는 Table 6에 제시하였다.

Table 6. Model fit statistics for the hierarchical logistic regression model of reading-while-listening condition

Step	χ^2	df	p	-2LL	Cox & Snell R^2	Nagelkerke R^2	Hosmer-Lemeshow test (p)
Step 1	6.473*	2	.039	177.267	.021	.046	.144
Step 2	28.075***	3	.000	155.665	.086	.193	.082
$\Delta\chi^2$	21.602***	1	.000				

* $p<.05$, *** $p<.001$

개별 변수의 회귀 계수를 분석한 결과, 연령의 회귀 계수가 유의하였으며($B=.047$, Wald=12.096, $p=.021$), 승산비는 1.048로 나타나 연령이 1단위(개월) 증가할 때 대상자가 VD로 분류될 승산이 약 4.8% 증가하는 것으로 나타났다. 또한 ‘듣기+읽기’ 조건 수행력의 회귀 계수가 유의하였으며($B=-.051$, Wald=18.524, $p=.000$), 승산비는 .950로, ‘듣기+읽기’ 조건 수행력이 1단위 증가할 때 대상자가 VD로 분류될 승산이 약 5.0% 감소하는 것으로 나타났다. 한편 KBIT2 점수는 2단계에서 통계적으로 유의하지 않아($p>.05$), 본 모형에서 집단 판별에 유의한 기여를 하지 않는 것으로 나타났다. ‘듣기+읽기’ 조건 수행력에 대한 로지스틱 회귀분석 결과는 Table 7에 제시하였다.

Table 7. Hierarchical logistic regression results for reading-while-listening condition

Predictor	B	SE	Wald	df	p	Exp(B)
Age	0.047**	0.014	12.096	1	.001	1.048
KBIT2	-0.022	0.018	1.420	1	.233	0.978
RWL accuracy	-0.051***	0.012	18.524	1	.000	0.950
Constant	-1.661	2.261	0.540	1	.463	0.190

Note. KBIT2=Korean Kaufman Brief Intelligence Test-second edition (Moon, 2020).

** $p<.01$, *** $p<.001$

로지스틱 회귀 모형의 분류 정확도를 확인한 결과, 전체 분류

정확도는 92.6%로 나타났으며, 집단별로 살펴보면 TD 집단은 100%가 정확하게 분류된 반면, VD 집단은 14.8%만이 정확하게 분류되었다(Table 8).

Table 8. Classification table for the logistic regression model of reading-while-listening condition

Observed	Predicted		Percentage correct (%)
	TD	VD	
TD	285	0	100.0
VD	23	4	14.8
Overall percentage	-	-	92.6

Note. TD=typically developing children; VD=children with vocabulary delay.

3. ‘읽기’ 조건 수행력의 판별 정확도

텍스트 제시 양식 중 ‘읽기’ 조건 수행력의 예측 요인을 확인하기 위해 2단계에서 ‘읽기’ 조건 수행력을 변수로 추가한 결과, 전체 모형은 통계적으로 유의하였으며($\chi^2=17.059$, $p=.001$), -2LL값은 177.267에서 166.681로 감소하여 모형 적합도가 향상되었다. 2단계에서 설명력이 Nagelkerke R^2 기준 .046에서 .120으로 증가하였으며, 따라서 투입 변수인 ‘읽기’ 조건 수행력이 모델 적합도를 유의하게 향상시키는 것으로 확인되었다($\Delta\chi^2=10.586$, $p=.001$). Hosmer-Lemeshow 검정 결과 역시 모형의 적합도가 양호한 것으로 나타났다($p>.05$). ‘읽기’ 조건 수행력을 투입한 위계적 로지스틱 회귀모형의 적합도 분석 결과는 Table 9에 제시하였다.

Table 9. Model fit statistics for the hierarchical logistic regression model of reading condition

Step	χ^2	df	p	-2LL	Cox & Snell R^2	Nagelkerke R^2	Hosmer-Lemeshow test (p)
Step 1	6.473*	2	.039	177.267	.021	.046	.144
Step 2	17.059**	3	.001	166.681	.053	.120	.285
$\Delta\chi^2$	10.586**	1	.001				

* $p<.05$, ** $p<.01$

개별 변수의 회귀 계수를 분석한 결과, 연령의 회귀 계수가 유의하였으며($B=.034$, Wald=7.983, $p=.005$), 승산비는 1.035로 나타나 연령이 1단위(개월) 증가할 때 대상자가 VD로 분류될 승산이 약 3.5% 증가하는 것으로 나타났다. 또한 ‘읽기’ 조건 수행력의 회귀 계수가 유의하였으며($B=-.032$, Wald=10.437, $p=.001$), 승산비는 .969로, ‘읽기’ 조건 수행력이 1단위 증가할 때 대상자가 VD로 분류될 승산이 약 3.1% 감소하는 것으로 나타났다. KBIT2 점수는 2단계에서 통계적으로 유의하지 않아($p>.05$), 본 모형에서 집단 판별에 유의한 기여를 하지 않는 것으로 나타났다.

‘읽기’ 조건 수행력에 대한 로지스틱 회귀분석 결과는 Table 10에 제시하였다.

Table 10. Hierarchical logistic regression results for reading condition

Predictor	<i>B</i>	<i>SE</i>	Wald	<i>df</i>	<i>p</i>	<i>Exp(B)</i>
Age	0.034**	0.012	7.983	1	.005	1.035
KBIT2	-0.025	0.017	2.043	1	.153	0.976
Reading accuracy	-0.032**	0.010	10.437	1	.001	0.969
Constant	-1.316	2.166	0.369	1	.544	0.268

Note. KBIT2=Korean Kaufman Brief Intelligence Test-second edition (Moon, 2020).

***p*<.01

로지스틱 회귀 모형의 분류 정확도를 확인한 결과, 전체 분류 정확도는 91.7%로 나타났으며, 집단별로 살펴보면 TD 집단은 100%가 정확하게 분류된 반면, VD 집단은 3.7%만이 정확하게 분류되었다(Table 11).

Table 11. Classification table for the logistic regression model of reading condition

Observed	Predicted		Percentage correct (%)
	TD	VD	
TD	285	0	100.0
VD	26	1	3.7
Overall percentage	-	-	91.7

Note. TD=typically developing children; VD=children with vocabulary delay.

4. ‘문학’ 지문 수행력의 판별 정확도

다음으로 지문 종류 중 ‘문학’ 지문 수행력의 예측 요인을 확인하기 위해 2단계에서 ‘문학’ 지문에 대한 수행력을 변수로 추가하여 모형의 변화를 검토하였으며, 그 결과 전체 모형은 통계적으로 유의하였고($\chi^2=25.870$, $p=.000$), -2LL값은 177.267에서 157.870으로 감소하여 모형 적합도가 향상되었다. 특히 2단계에서 설명력이 Nagelkerke R^2 기준 .046에서 .179로 증가하였으며, 따라서 투입 변수인 ‘문학’ 지문 수행력이 모델 적합도를 유의하게 향상시키는 것으로 확인되었다($\Delta\chi^2=19.397$, $p=.000$). Hosmer-Lemeshow 검정 결과 역시 모형의 적합도가 양호한 것으로 나타났다($p>.05$). ‘문학’ 지문 수행력을 투입한 위계적 로지스틱 회귀모형의 적합도 분석 결과는 Table 12에 제시하였다.

Table 12. Model fit statistics for the hierarchical logistic regression model of fiction condition

Step	χ^2	<i>df</i>	<i>p</i>	-2LL	Cox & Snell R^2	Nagelkerke R^2	Hosmer-Lemeshow test (<i>p</i>)
Step 1	6.473*	2	.039	177.267	.021	.046	.144
Step 2	25.870***	3	.000	157.870	.080	.179	.418
$\Delta\chi^2$	19.397***	1	.000				

p*<.05, **p*<.001

개별 변수의 회귀 계수를 분석한 결과, 연령의 회귀 계수가 유의하였으며($B=.040$, Wald=10.031, $p=.002$), 승산비는 1.041로 나타났다. 이는 연령이 1단위(개월) 증가할 때 대상자가 VD로 분류될 승산이 약 4.1% 증가함을 의미한다. 또한 ‘문학’ 지문 수행력의 회귀 계수가 유의하였으며($B=-.066$, Wald=17.261, $p=.000$), 승산비는 .936으로, ‘문학’ 지문 수행력이 1단위 증가할 때 대상자가 VD로 분류될 승산이 약 6.4% 감소하는 것으로 나타났다. 한편 KBIT2 점수는 2단계에서 통계적으로 유의하지 않아($p>.05$), 본 모형에서 집단 판별에 유의한 기여를 하지 않는 것으로 나타났다. ‘문학’ 지문 수행력에 대한 로지스틱 회귀분석 결과는 Table 13에 제시하였다.

Table 13. Hierarchical logistic regression results for fiction condition

Predictor	<i>B</i>	<i>SE</i>	Wald	<i>df</i>	<i>p</i>	<i>Exp(B)</i>
Age	0.040**	0.013	10.031	1	.002	1.041
KBIT2	-0.021	0.018	1.421	1	.233	0.979
Fiction accuracy	-0.066***	0.016	17.261	1	.000	0.936
Constant	0.768	2.108	0.133	1	.716	2.155

Note. KBIT2=Korean Kaufman Brief Intelligence Test-second edition (Moon, 2020).

p*<.01, *p*<.001

로지스틱 회귀 모형의 분류 정확도를 확인한 결과, 전체 분류 정확도는 92.0%로 나타났으며, 집단별로 살펴보면 TD 집단은 100%가 정확하게 분류된 반면, VD 집단은 7.4%만이 정확하게 분류되었다(Table 14).

Table 14. Classification table for the logistic regression model of fiction condition

Observed	Predicted		Percentage correct (%)
	TD	VD	
TD	285	0	100.0
VD	25	2	7.4
Overall percentage	-	-	92.0

Note. TD=typically developing children; VD=children with vocabulary delay.

5. ‘비문학’ 지문 수행력의 판별 정확도

지문 종류 중 ‘비문학’ 지문 수행력의 예측 요인을 확인하기 위해 2단계에서 ‘비문학’ 지문 수행력을 변수로 추가하여 살펴본 결과, 전체 모형은 통계적으로 유의하였고($\chi^2=29.557$, $p=.000$), -2LL값은 177.267에서 154.183으로 감소하여 모형 적합도가 향상되었다. 2단계에서 설명력이 Nagelkerke R^2 기준 .046에서 .203으로 증가하였으며, 따라서 투입 변수인 ‘비문학’ 지문 수행력이 모형 적합도를 유의하게 향상시키는 것으로 확인되었다($\Delta\chi^2=23.084$, $p=.000$). Hosmer-Lemeshow 검정 결과 역시 모형의 적합도가 양호한 것으로 나타났다($p>.05$). ‘비문학’ 지문 수행력을 투입한 위계적 로지스틱 회귀모형의 적합도 분석 결과는 Table 15에 제시하였다.

Table 15. Model fit statistics for the hierarchical logistic regression model of expository condition

Step	χ^2	df	p	-2LL	Cox & Snell R^2	Nagelkerke R^2	Hosmer-Lemeshow test (p)
Step 1	6.473*	2	.039	177.267	.021	.046	.144
Step 2	29.557***	3	.000	154.183	.090	.203	.587
$\Delta\chi^2$	23.084***	1	.000				

* $p<.05$, *** $p<.001$

개별 변수의 회귀 계수를 분석한 결과, 연령의 회귀 계수가 유의하였으며($B=.060$, Wald=15.502, $p=.000$), 승산비는 1.062로 나타나 연령이 1단위(개월) 증가할 때 대상자가 VD로 분류될 승산이 약 6.2% 증가하는 것으로 나타났다. 또한 ‘비문학’ 지문 수행력의 회귀 계수가 유의하였으며($B=-.060$, Wald=18.910, $p=.000$), 승산비는 .941로, ‘비문학’ 지문 수행력이 1단위 증가할 때 대상자가 VD로 분류될 승산이 약 5.9% 감소하는 것으로 나타났다. 한편 KBIT2 점수는 2단계에서 통계적으로 유의하지 않아($p>.05$), 본 모형에서 집단 판별에 유의한 기여를 하지 않는 것으로 나타났다. ‘비문학’ 지문 수행력에 대한 로지스틱 회귀분석 결과는 Table 16에 제시하였다.

Table 16. Hierarchical logistic regression results for expository condition

Predictor	B	SE	Wald	df	p	Exp(B)
Age	0.060***	0.015	15.502	1	.000	1.062
KBIT2	-0.016	0.018	0.754	1	.385	0.984
Expository accuracy	-0.060***	0.014	18.910	1	.000	0.941
Constant	-3.780	2.366	2.552	1	.110	0.023

Note. KBIT2=Korean Kaufman Brief Intelligence Test-second edition (Moon, 2020).

*** $p<.001$

로지스틱 회귀 모형의 분류 정확도를 확인한 결과, 전체 분류

정확도는 92.6%로 나타났으며, 집단별로 살펴보면 TD 집단은 100%가 정확하게 분류된 반면, VD 집단은 14.8%만이 정확하게 분류되었다(Table 17).

Table 17. Classification table for the logistic regression model of expository condition

Observed	Predicted		Percentage correct (%)
	TD	VD	
TD	285	0	100.0
VD	23	4	14.8
Overall percentage	-	-	92.6

Note. TD=typically developing children; VD=children with vocabulary delay.

6. ‘사실적 이해’ 문항 수행력의 판별 정확도

질문 유형 중 ‘사실적 이해’ 질문에 대한 수행력의 예측 요인을 확인하기 위해 2단계에서 ‘사실적 이해’ 수행력을 변수로 추가하여 모형의 변화를 검토한 결과, 전체 모형은 통계적으로 유의하였으며($\chi^2=27.134$, $p=.000$), -2LL값은 177.267에서 156.606으로 감소하여 모형 적합도가 향상되었다. 2단계에서 설명력이 Nagelkerke R^2 기준 .046에서 .187로 증가하였으며, 따라서 투입 변수인 ‘사실적 이해’ 질문 수행력이 모형 적합도를 유의하게 향상시키는 것으로 확인되었다($\Delta\chi^2=20.661$, $p=.000$). Hosmer-Lemeshow 검정 결과 역시 모형의 적합도가 양호한 것으로 나타났다($p>.05$). ‘사실적 이해’ 문항 수행력을 투입한 위계적 로지스틱 회귀모형의 적합도 분석 결과는 Table 18에 제시하였다.

Table 18. Model fit statistics for the hierarchical logistic regression model of literal condition

Step	χ^2	df	p	-2LL	Cox & Snell R^2	Nagelkerke R^2	Hosmer-Lemeshow test (p)
Step 1	6.473*	2	.039	177.267	.021	.046	.144
Step 2	27.134***	3	.000	156.606	.083	.187	.628
$\Delta\chi^2$	20.661***	1	.000				

* $p<.05$, *** $p<.001$

개별 변수의 회귀 계수를 분석한 결과, 연령의 회귀 계수가 유의하였으며($B=.047$, Wald=11.977, $p=.000$), 승산비는 1.048로 나타나 연령이 1단위(개월) 증가할 때 대상자가 VD로 분류될 승산이 약 4.8% 증가하는 것으로 나타났다. 또한 ‘사실적 이해’ 수행력의 회귀 계수가 유의하였으며($B=-.067$, Wald=17.452, $p=.000$), 승산비는 .936으로, ‘사실적 이해’ 문항 수행력이 1단위 증가할 때 대상자가 VD로 분류될 승산이 약 6.4% 감소하는 것으로 나타났다. 한편 KBIT2 점수는 2단계에서 통계적으로 유의하지 않아($p>.05$), 본 모형에서 집단 판별에 유의한 기여

를 하지 않는 것으로 나타났다. ‘사실적 이해’ 문항 수행력에 대한 로지스틱 회귀분석 결과는 Table 19에 제시하였다.

Table 19. Hierarchical logistic regression results for literal condition

Predictor	<i>B</i>	<i>SE</i>	Wald	<i>df</i>	<i>p</i>	<i>Exp(B)</i>
Age	0.047**	0.014	11.977	1	.001	1.048
KBIT2	-0.017	0.018	0.909	1	.340	0.983
Literal accuracy	-0.067***	0.016	17.452	1	.000	0.936
Constant	-0.914	2.121	0.186	1	.666	0.401

Note. KBIT2=Korean Kaufman Brief Intelligence Test-second edition (Moon, 2020).

p*<.01, *p*<.001

로지스틱 회귀 모형의 분류 정확도를 확인한 결과, 전체 분류 정확도는 92.0%로 나타났으며, 집단별로 살펴보면 TD 집단은 100%가 정확하게 분류된 반면, VD 집단은 7.4%만이 정확하게 분류되었다(Table 20).

Table 20. Classification table for the logistic regression model of literal condition

Observed	Predicted		Percentage correct (%)
	TD	VD	
TD	285	0	100.0
VD	25	2	7.4
Overall percentage	-	-	92.0

Note. TD=typically developing children; VD=children with vocabulary delay.

7. ‘추론적 이해’ 문항 수행력의 판별 정확도

마지막으로 질문 유형 중 ‘추론적 이해’ 질문에 대한 수행력의 예측 요인을 확인하기 위해 2단계에서 ‘추론적 이해’ 수행력을 변수로 추가한 결과, 전체 모형은 통계적으로 유의하였으며($\chi^2=26.048$, $p=.000$), -2LL값은 177.267에서 157.692로 감소하여 모형 적합도가 향상되었다. 2단계에서 설명력이 Nagelkerke R^2 기준 .046에서 .180으로 증가하였으며, 따라서 투입 변수인 ‘추론적 이해’ 수행력이 모델 적합도를 유의하게 향상시키는 것으로 확인되었다($\Delta\chi^2=19.575$, $p=.000$). 한편 Hosmer-Lemeshow 검정 결과가 통계적으로 유의하여($p=.024$) 모형의 적합도에 일부 한계가 있는 것으로 나타났다. 다만, Hosmer-Lemeshow 검정은 표본의 특성이나 집단 비율의 불균형에 영향을 받을 수 있다는 점이 보고된 바(Hosmer et al., 2013; Kramer & Zimmerman, 2007), 본 결과는 신중하게 해석할 필요가 있다. ‘추론적 이해’ 문항 수행력을 투입한 위계적 로지스틱 회귀모형의 적합도 분석 결과는 Table 21에 제시하였다.

Table 21. Model fit statistics for the hierarchical logistic regression model of inferential condition

Step	χ^2	<i>df</i>	<i>p</i>	-2LL	Cox & Snell R^2	Nagelkerke R^2	Hosmer-Lemeshow test (<i>p</i>)
Step 1	6.473*	2	.039	177.267	.021	.046	.144
Step 2	26.048***	3	.000	157.692	.080	.180	.024
$\Delta\chi^2$	19.575***	1	.000				

p*<.05, **p*<.001

개별 변수의 회귀 계수를 분석한 결과, 연령의 회귀 계수가 유의하였으며($B=.053$, Wald=13.086, $p=.000$), 승산비는 1.055로 나타나 연령이 1단위(개월) 증가할 때 대상자가 VD로 분류될 승산이 약 5.5% 증가하는 것으로 나타났다. 또한 ‘추론적 이해’ 수행력의 회귀 계수가 유의하였으며($B=-.061$, Wald=16.881, $p=.000$), 승산비는 .941로, ‘추론적 이해’ 문항 수행력이 1단위 증가할 때 대상자가 VD로 분류될 승산이 약 5.9% 감소하는 것으로 나타났다. 한편 KBIT2 점수는 2단계에서 통계적으로 유의하지 않아($p>.05$), 본 모형에서 집단 판별에 유의한 기여를 하지 않는 것으로 나타났다. ‘추론적 이해’ 문항 수행력에 대한 로지스틱 회귀분석 결과는 Table 22에 제시하였다.

Table 22. Hierarchical logistic regression results for inferential condition

Predictor	<i>B</i>	<i>SE</i>	Wald	<i>df</i>	<i>p</i>	<i>Exp(B)</i>
Age	0.053***	0.015	13.086	1	.000	1.055
KBIT2	-0.020	0.017	1.323	1	.250	0.980
Inferential accuracy	-0.061***	0.015	16.881	1	.000	0.941
Constant	-1.946	2.180	0.797	1	.372	0.143

Note. KBIT2=Korean Kaufman Brief Intelligence Test-second edition (Moon, 2020).

****p*<.001

로지스틱 회귀 모형의 분류 정확도를 확인한 결과, 전체 분류 정확도는 91.3%로 나타났으며, 집단별로 살펴보면 TD 집단은 99.6%가 정확하게 분류된 반면, VD 집단은 3.7%만이 정확하게 분류되었다(Table 23).

Table 23. Classification table for the logistic regression model of inferential condition

Observed	Predicted		Percentage correct (%)
	TD	VD	
TD	284	1	99.6
VD	26	1	3.7
Overall percentage	-	-	91.3

Note. TD=typically developing children; VD=children with vocabulary delay.

8. ROC 곡선 분석을 통한 모형의 판별력 비교

본 연구에서는 이분형 종속변수인 집단 판별력을 예측하기 위하여 먼저 위계적 로지스틱 회귀분석을 통해 여러 예측 모형을 구축한 후, 각 모형의 판별력을 비교하기 위하여 ROC(Receiver Operating Characteristic) 곡선 분석 및 AUC(Area Under the Curve) 비교를 추가로 수행하였다. 로지스틱 회귀 분석 및 예측 확률 산출은 SPSS 25.0을 사용하였으며, SPSS에서 산출된 예측 확률값을 R(4.3.0)의 pROC 패키지를 활용하여 ROC 곡선을 도출하고 AUC 값을 산출하였다.

먼저 7개 변인들의 판별력을 시각적으로 비교하기 위하여 ROC 곡선을 산출하였다(Figure 1). ROC 곡선의 전체적인 분포를 살펴본 결과, 모든 예측 모형의 ROC 곡선이 무작위 분류 기준선의 좌측 상단에 위치하여, 각 모형이 TD 아동과 VD 아동을 구분하는데에 일정 수준 이상의 판별력을 갖고 있음을 확인하였다.

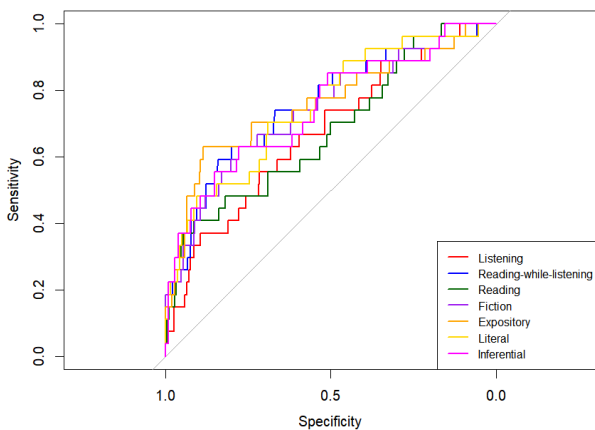


Figure 1. ROC Curves for text comprehension classification models

또한 7개 변인의 예측 확률을 사용하여 AUC 분석을 실시하였다. 예측 모형별 AUC 분석 결과는 Table 24에 제시하였다. 전체 7개 모형의 AUC는 .674에서 .759 사이였으며, AUC의 95% 신뢰구간은 각 모형별로 .61~.81 범위에 해당하였다. 분석 결과 ‘듣기+읽기’ 조건 기반 모형(AUC=.758), ‘비문학 지문’ 기반 모형(AUC=.758), 그리고 ‘사실적 이해’ 기반 모형(AUC=.750)이 상대적으로 높은 판별력을 보였다. 반면, ‘듣기’ 조건 기반 모형(AUC=.674)과 ‘읽기’ 조건 기반 모형(AUC=.678)은 AUC값이 .70 미만으로 나타나, 다른 모형들에 비해 상대적으로 낮은 판별력을 가진 것으로 나타났다.

ROC 분석 결과를 기반으로 각 예측 모형의 분류 성능을 보다 구체적으로 확인하기 위해 Youden Index가 최대가 되는 지점에서 최적 절단값(threshold)을 산출하고, 해당 기준에서의 민감도(sensitivity)와 특이도(specificity)를 산출하였다(Table 25). 분석 결과, ‘사실적 이해’ 모형은 가장 낮은 절단값(.078)에서 최대의 민감도(.704)를 보여 VD 아동을 가장 민감하게 탐지하는 모형으로 나타났다. 그러나 이 모형의 특이도는 상대적으로 낮아(.691) TD 아동을 VD 아동으로 오분류할 가능성이 다른 모형에 비해 높았다. 또한 ‘읽기’ 조건 모형은 절단값이 가장 높게 산출되었으며

(.145), 민감도는 가장 낮았지만(.407) 특이도는 가장 높아(.916) TD 아동을 가장 정확하게 판별하는 것으로 나타났다. 한편 비문학 지문(절단값=.146), 문학 지문(절단값=.104), 추론적 이해(절단값=.106) 모형은 모두 중간 수준의 절단값을 보였으며, 세 모형 모두 민감도(.630)와 특이도(.779~.888)는 비교적 균형 있게 나타나 안정적인 분류 성능을 보였다. 마지막으로 듣기(절단값=.098, 민감도=.556, 특이도=.716) 및 듣기+읽기(절단값=.120, 민감도=.593, 특이도=.842) 모형은 특이도는 비교적 높으나 민감도는 중간 수준으로 나타나, 전체적으로 중간 정도의 판별력을 가진 것으로 나타났다.

Table 24. AUC results of text comprehension classification models

Predictor	AUC	95% CI ^a	
		LL ^b	UL ^c
Listening comprehension	.674	.566	.781
RWL comprehension	.758	.658	.858
Reading comprehension	.678	.567	.790
Fiction comprehension	.743	.640	.847
Expository comprehension	.758	.647	.870
Literal comprehension	.750	.651	.850
Inferential comprehension	.745	.639	.850

Note. RWL=reading-while-listening.

^a CI=confidence interval. ^b LL=lower limit. ^c UL=upper limit.

Table 25. Cutoff values, sensitivity, and specificity based on the Youden index for each predictor

Predictor	Cutoff value	Sensitivity	Specificity
Listening comprehension	.098	.556	.716
RWL comprehension	.120	.593	.842
Reading comprehension	.145	.407	.916
Fiction comprehension	.104	.630	.779
Expository comprehension	.146	.630	.888
Literal comprehension	.078	.704	.691
Inferential comprehension	.106	.630	.779

Note. RWL=reading-while-listening.

IV. 논의 및 결론

본 연구는 초등학교 1~6학년 일반 아동 집단과 어휘발달 지연 아동 집단을 유의하게 판별하는 텍스트 이해 검사 조건이 무엇인지 검토하고자 하였다. 이를 위해 텍스트 제시 양식 3개(듣기, 듣기+읽기, 읽기), 지문 종류 2개(문학 지문, 비문학 지문), 질문 유형 2개(사실적 이해, 추론적 이해)를 독립변수로 구성하였으며, 이와 같은 7개의 세부 텍스트 과제 조건을 각각 독립변수로 투입하

여 위계적 로지스틱 회귀분석(hierarchical logistic regression)을 실시하였다. 연령과 지능의 영향을 통제하기 위하여 모든 분석에서 연구 대상 아동들의 월령과 KBIT2 점수를 공변량으로 투입함으로써, 각 텍스트 과제 조건이 집단의 판별에 미치는 고유한 기여도를 측정하였다.

7개의 텍스트 과제 조건 각각을 독립변수로 투입하여 로지스틱 회귀분석을 실시한 결과, 투입된 변수들은 모두 모형의 설명력을 유의하게 증가시키는 것으로 나타났다. 이러한 결과는 각각의 변수가 TD 아동과 VD 아동을 구분하는 데 유의하게 기여함을 시사한다. 그런데 로지스틱 회귀 분석의 분류표를 살펴보면 TD 아동들에 대한 분류 정확도가 추론적 이해 지문 수행력이 99.6%였던 것을 제외하고 다른 모든 모형에서 100%로 나타났으나, VD 아동들에 대한 분류 정확도는 3.7%~14.8% 사이에 머물렀다. 이는 분석의 대상이 된 두 집단의 표본 크기에 상당한 불균형이 존재하며, VD 집단의 경우 TD 집단에 비해 수행력의 변동성이 더 높아, 로지스틱 회귀분석과 같은 단일 절단점(= .50) 기준에서 VD 집단을 안정적으로 판별하기 어려웠던 점을 그 원인으로 해석할 수 있다. 이러한 연구 결과는 기본 절단값(.50)을 적용할 때 로지스틱 회귀 모형이 소수인 집단을 과소 분류하는 경향이 있음을 보고한 선행 연구들(Fawcett, 2006; He & Garcia, 2009; King & Zeng, 2001)과도 일치하는 결과이다.

이 외에도 VD 집단의 이질성도 분류 정확도에 일부 영향을 미쳤을 가능성이 있다. 본 연구에서는 REVT의 수용어휘력 또는 표현어휘력 검사 결과 10%ile 미만인 아동을 VD 아동으로 선별하였는데, 이 기준은 어휘 능력의 취약성을 확인하는 데에는 유용하지만, 수용어휘와 표현어휘의 수행 양상이 서로 다른 아동들을 동일 집단 내에 포함시킬 수 있다. 또한 어휘 능력에 제한을 보이는 아동들 가운데에도 다른 언어 능력 및 인지적 특성에는 개인차가 존재할 수 있으므로, 집단 내부의 변동성이 증가했을 가능성이 있다. 이러한 집단 내 이질성은 하나의 절단값만으로 집단을 안정적으로 구분하는 데에 한계를 초래했을 가능성이 있으며, 결과적으로 분류 정확도가 낮게 나타난 요인 중 하나로 해석할 수 있다.

로지스틱 회귀분석을 통해 7개 변수들이 각각 집단 차이를 설명함을 확인한 바, 각 변수들의 판별력을 구체적으로 비교하기 위하여 ROC 곡선 분석 및 AUC 분석을 실시하였다. 분석 결과 7개 예측 모형의 ROC 곡선은 무작위 분류 기준선 상단에 위치하여 일정 수준 이상의 분류 정확도를 갖고 있음을 확인하였으며, AUC 값을 비교한 결과 듣기+읽기, 비문학 지문, 사실적 이해 세 가지 수행 조건에 기반한 모형이 상대적으로 높은 판별력을 가진 것으로 나타났다. 특히 Youden Index 분석 결과를 검토해 보면, 사실적 이해 모형은 VD 아동의 판별에 가장 높은 민감도를 보였으며(.704), 읽기 조건 모형은 특이도가 가장 높아(.916) TD 아동을 VD 아동으로 오분류할 가능성이 가장 낮은 모형으로 확인되었다. 또한 비문학 지문 이해 모형은 민감도(.630)와 특이도(.888)가 균형 있게 산출되어 안정적인 분류 성능을 보이는 모형으로 나타났다.

종합하면, 텍스트 이해 과제의 특징에 따라 집단에 대한 판별 우위와 임상적 유용성이 서로 다르게 나타났다. VD 아동을 민감하게 판별하는 데에는 ‘사실적 이해’ 과제가 가장 효과적이었으며,

TD 아동을 VD 아동으로 오분류하는 것을 최소화하는 데에는 ‘읽기’ 과제가 가장 우수한 성능을 보였다. 또한 ‘비문학’ 지문 이해 과제는 민감도와 특이도의 균형 측면에서 가장 안정적인 판별력을 나타냈다. 이러한 결과는 텍스트 이해 평가에서 하나의 과제나 단일 총점을 활용하여 아동의 수행력을 판단하는 방식에 한계가 있음을 시사한다. 이를 고려하면 평가의 목적이 선별인지, 진단인지, 또는 중재 계획 수립인지를 고려하여 과제의 제시 양식과 지문 종류, 질문 유형을 전략적으로 선택할 필요가 있다. 예를 들어 언어발달지연이 의심되는 위험군 아동을 놓치지 않고 탐지하는 것이 목표인 경우에는 사실적 이해 과제를 중심으로 평가를 실시할 수 있으며, 일반 아동을 과잉 진단하는 것을 최소화해야 하는 상황에서는 읽기 과제를 활용하는 것이 유용할 수 있다. 또한 비문학 지문은 민감도와 특이도가 균형적인 과제인 것으로 나타난 바, 임상 현장에서 안정적인 평가 지표로 활용될 수 있을 것이다.

특히 본 연구의 주요 대상인 VD 아동을 판별하는 데 있어 사실적 이해 과제가 가장 민감한 것으로 나타난 결과는, 연구 대상이 REVT를 기준으로 어휘 능력에 제한을 보이는 아동으로 선정되었다는 점과 관련이 있는 것으로 해석된다. 즉, 본 연구의 대상인 VD 아동의 핵심적인 특성은 단어 의미 지식의 폭과 깊이가 제한되어 있다는 점인데, 본 연구의 결과는 이러한 어휘 능력의 제한이 사실적 이해 과제의 수행에 보다 민감하게 반영되었음을 시사한다. 사실적 이해는 텍스트에 명시적으로 제시된 정보를 정확하게 이해하고 회상하는 과정이 필요한 과제이므로, 어휘 지식에 제한이 있는 아동의 경우 텍스트 이해의 초기 단계에서 어려움이 나타날 수 있다. 이러한 점을 고려할 때 본 연구의 대상인 VD 아동은 텍스트의 명시적 정보를 처리하는 단계에서부터 수행 저하가 나타날 수 있다. 반면 추론적 이해는 텍스트에 명시되지 않은 정보를 배경 지식, 문맥 단서, 정보 간 통합을 바탕으로 해석해야 하므로, 어휘 능력뿐 아니라 고차원적 언어 처리 능력이 함께 요구되는 과제이다. 이와 같은 특성으로 인해 TD 아동에게서도 수행 난이도가 높아져 두 집단 간 수행 차이가 상대적으로 감소하였고, 결과적으로 추론적 이해 과제의 집단 판별력이 사실적 이해 과제보다 낮게 나타난 것으로 해석할 수 있다.

본 연구의 의의는 특정 텍스트 이해 과제가 다른 텍스트 이해 과제들에 비해 우수하다는 점을 밝힌 데에 있기보다는, 텍스트 이해 수행력을 구성하는 세부 조건들이 서로 다른 임상적 정보를 제공한다는 점을 실증적으로 확인한 데에 있다. 지금까지 텍스트 이해 평가는 주로 총점이나 평균 점수와 같은 수행력을 기준으로 해석되어 왔으나, 본 연구 결과는 과제의 제시 양식, 지문의 특성, 질문 유형에 따라 집단 판별에 기여하는 정도가 달라질 수 있음을 보여준다. 이는 텍스트 이해 능력이 단일한 능력으로 설명되기보다, 여러 하위 요소들의 상호작용을 통해 나타나는 복합적인 구성 개념임을 시사한다. 따라서 향후 텍스트 이해 평가는 단일 점수를 산출하는 수준에서 나아가, 아동이 어떠한 텍스트 조건에서 강점과 약점을 보이는지 다차원적으로 분석하는 방향으로 발전할 필요가 있을 것이다. 이러한 접근은 언어발달에 어려움을 보이는 아동의 특성을 보다 정교하게 파악할 수 있게 하며, 평가 결과를 중재 계획과 직접적으로 연결하는 데에도 기여할 수 있을 것으로 기대된다.

한편, 본 연구는 서울 소재 초등학교에 재학 중인 초등학교생 연구 대상으로 하여 표집의 지역적 대표성에 제한이 있다. 또한 일반 아동 대비 어휘발달 지연 아동의 표본 크기가 상대적으로 작아, 이러한 표본 불균형이 로지스틱 회귀분석의 분류 결과에 영향을 미쳤을 가능성을 배제할 수 없다. 따라서 후속 연구에서는 다양한 지역에서 표본을 균형 있게 확보하고, 아동의 언어 능력, 읽기 수준, 배경지식 등이 충분히 대표성을 갖도록 표집 설계가 이루어져야 할 것이다. 또한 본 연구는 통제된 실험 목적에 따라 어휘발달 지연 아동을 주요 연구 대상으로 설정하였으나, 아동의 전반적인 언어 수행력이 어휘 능력만으로 온전히 설명되지는 않는다는 한계가 있다. 따라서 다양한 언어 및 발달 특성을 가진 임상 집단을 포함하여 본 연구에서 도출된 다양한 텍스트 이해 과제 조건별 판별 유효성이 다양한 집단에서도 일관되게 나타나는지 검증할 필요가 있다. 마지막으로 본 연구에서 사용한 문학 지문과 비문학 지문은 지문을 구성하는 문장 수에 차이가 있었으므로, 지문 종류 외에도 텍스트 길이와 정보량 차이가 일부 영향을 미쳤을 가능성이 있다. 또한 평균어절길이는 유사하게 통제하였으나, 어휘 다양도, 주제 친숙도 및 배경지식과 같은 텍스트 특성은 충분히 반영하지 못하였다. 따라서 후속 연구에서는 문장 수, 총 어절 수, 정보 밀도뿐 아니라 어휘적 특성, 주제 친숙도, 배경지식 등을 보다 종합적으로 고려한 텍스트를 활용하여 본 연구의 결과를 재검증할 필요가 있다.

Reference

- Bae, M. Y. (2025). *The effects of reading media types and reading amount on text comprehension: Focusing on fifth-grade elementary school students* (Master's thesis). Seoul National University of Education, Seoul.
- Best, R. M., Floyd, R. G., & McNamara, D. S. (2008). Differential competencies contributing to children's comprehension of narrative and expository texts. *Reading Psychology, 29*(2), 137-164. doi:10.1080/02702710801963951
- Biemiller, A. (2003). Vocabulary: Needed if more children are to read well. *Reading Psychology, 24*(3-4), 323-335. doi:10.1080/02702710390227297
- Bishop, D. V. M. (2014). *Uncommon understanding: Development and disorders of language comprehension in children* (Classic ed.). Psychology Press.
- Bishop, D. V. M., & Snowling, M. J. (2004). Developmental dyslexia and specific language impairment: Same or different? *Psychological Bulletin, 130*(6), 858-886. doi:10.1037/0033-2909.130.6.858
- Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & Klee, T. M. (2017). CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development. Phase 2. *Journal of Child Psychology and Psychiatry, 58*(10), 1068-1080. doi:10.1111/jcpp.12721
- Cain, K., & Oakhill, J. (2006). Profiles of children with specific reading comprehension difficulties. *British Journal of Educational Psychology, 76*(4), 683-696. doi:10.1348/000709905x67610
- Cain, K., & Oakhill, J. (2008). *Children's comprehension problems in oral and written language: A cognitive perspective*. New York: Guilford Press.
- Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology, 96*(1), 31-42. doi:10.1037/0022-0663.96.1.31
- Catts, H. W., Hogan, T. P., & Fey, M. E. (2003). Subgrouping poor readers on the basis of individual differences in reading-related abilities. *Journal of Learning Disabilities, 36*(2), 151-164. doi:10.1177/002221940303600208
- Chung, H., Jeong, E., Nam, S., Kim, J., Yoon, S. R., Kim, A., & Yim, D. (2023b). Story comprehension skills in school-aged children according to text type and story modality. *Korean Journal of Child Studies, 44*(4), 389-400. doi:10.5723/kjcs.2023.44.4.389
- Chung, H., Song, H., Cho, Y., Oh, Y., Joo, H., & Yim, D. (2023a). Story comprehension skills of school-aged children by passage and question type according to story conditions. *Journal of Speech-Language & Hearing Disorders, 32*(2), 49-60. doi:10.15724/jslhd.2023.32.2.049
- Conti-Ramsden, G., & Durkin, K. (2016). What factors influence language impairment considering resilience as well as risk. *Folia Phoniatrica et Logopaedica, 67*(6), 293-299. doi:10.1159/000444750
- DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). Sage.
- Duke, N. K., & Roberts, K. L. (2010). The genre-specific nature of reading comprehension. In D. Wyse, R. Andrews, & J. Hoffman (Eds.), *The Routledge international handbook of English, language and literacy teaching* (pp. 74-83). Routledge. doi:10.4324/9780203863091.ch7
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861-874. doi:10.1016/j.patrec.2005.10.010
- Goldman, S. R., & Rakestraw, J. A., Jr. (2000). Structural aspects of constructing meaning from text. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 311-356). Lawrence Erlbaum Associates.
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*(1), 6-10. doi:10.1177/074193258600700104
- Graesser, A. C., McNamara, D. S., & Louwerse, M. M. (2003). What do readers need to learn in order to process coherence relations in narrative and expository text? In A. P. Sweet & C. E. Snow (Eds.), *Rethinking reading comprehension* (pp. 82-98). New York: Guilford Press.
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*(3), 371-395. doi:10.1037/0033-295x.101.3.371
- Gray, S. (2004). Word learning by preschoolers with specific language

- impairment. *Journal of Speech, Language, and Hearing Research*, 47(5), 1117-1132. doi:10.1044/1092-4388(2004/083)
- He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263-1284. doi:10.1109/TKDE.2008.239
- Hoffman, L. M., & Gillam, R. B. (2004). Verbal and spatial information processing constraints in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 47(1), 114-125. doi:10.1044/1092-4388(2004/011)
- Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, 2, 127-160. doi:10.1007/bf00401799
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Hoboken, NJ: Wiley.
- Kendeou, P., van den Broek, P., White, M. J., & Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. *Journal of Educational Psychology*, 101(4), 765-778. doi:10.1037/a0015956
- Kim, Y. T., Hong, G. H., Kim, K. H., Jang, H. S., & Lee, J. Y. (2009). *Receptive & Expressive Vocabulary Test* (REVT). Seoul: Seoul Community Rehabilitation Center.
- King, G., & Zeng, L. (2001). Logistic regression in rare events data. *Political Analysis*, 9(2), 137-163. doi:10.1093/oxfordjournals.pan.a004868
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge: Cambridge University Press.
- Kramer, A. A., & Zimmerman, J. E. (2007). Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. *Critical Care Medicine*, 35(9), 2052-2056. doi:10.1097/01.ccm.0000275267.64078.b0
- Leclercq, A.-L., Majerus, S., Prigent, G., & Maillart, C. (2013). The impact of dual tasking on sentence comprehension in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 56(1), 265-280. doi:10.1044/1092-4388(2012/10-0290)
- Leonard, L. B., & Schroeder, M. L. (2024). The study of children with developmental language disorder beyond English: A tutorial. *Language Acquisition*, 31(3-4), 180-198. doi:10.1080/10489223.2023.2197891
- McGregor, K. K., Oleson, J., Bahnsen, A., & Duff, D. (2013). Children with developmental language impairment have vocabulary deficits characterized by limited breadth and depth. *International Journal of Language and Communication Disorders*, 48(3), 307-319. doi:10.1111/1460-6984.12008
- Meyer, B. J. F., & Ray, M. N. (2011). Structure strategy interventions: Increasing reading comprehension of expository text. *International Electronic Journal of Elementary Education*, 4(1), 127-152.
- Montali, J., & Lewandowski, L. (1996). Bimodal reading: Benefits of a talking computer for average and less skilled readers. *Journal of Learning Disabilities*, 29(3), 271-279. doi:10.1177/002221949602900305
- Moon, S. B. (2020). *Korean Kaufman Brief Intelligence Test-second edition (KBIT2)*. Seoul: Inpsyt Psychological Test Institute.
- Nation, K. (2019). Children's reading difficulties, language, and reflections on the simple view of reading. *Australian Journal of Learning Difficulties*, 24(1), 47-73. doi:10.1080/19404158.2019.1609272
- Nation, K., & Snowling, M. J. (1998). Semantic processing and the development of word-recognition skills: Evidence from children with reading comprehension difficulties. *Journal of Memory and Language*, 39(1), 85-101. doi:10.1006/jmla.1998.2564
- Nation, P. (2005). Teaching and learning vocabulary. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 581-595). Routledge.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.
- Oakhill, J., & Cain, K. (2012). The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading*, 16(2), 91-121. doi:10.1080/10888438.2010.529219
- Oakhill, J., Cain, K., & Elbro, C. (2014). *Understanding and teaching reading comprehension: A handbook*. Routledge. doi:10.4324/9781315756042
- Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, 98(3), 554-566. doi:10.1037/0022-0663.98.3.554
- Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11(4), 357-383. doi:10.1080/10888430701530730
- Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, 18(1), 22-37. doi:10.1080/10888438.2013.827687
- Quinn, J. M., Wagner, R. K., Petscher, Y., & Lopez, D. (2015). Developmental relations between vocabulary knowledge and reading comprehension: A latent change score modeling study. *Child Development*, 86(1), 159-175. doi:10.1111/cdev.12292
- Snow, C. E. (2010). Academic language and the challenge of reading for learning about science. *Science*, 328(5977), 450-452. doi:10.1126/science.1182597
- Stothard, S. E., Snowling, M. J., Bishop, D. V. M., Chipchase, B. B., & Kaplan, C. A. (1998). Language-impaired preschoolers: A follow-up into adolescence. *Journal of Speech, Language, and Hearing Research*, 41(2), 407-418. doi:10.1044/jslhr.4102.407
- Valentini, A., Ricketts, J., Pye, R. E., & Houston-Price, C. (2018). Listening while reading promotes word learning from stories. *Journal of Experimental Child Psychology*, 167, 10-31. doi:10.1016/j.jecp.2017.09.022
- Verhoeven, L., & Perfetti, C. (2008). Advances in text comprehension: Model, process and development. *Applied Cognitive Psychology*, 22(3), 293-301. doi:10.1002/acp.1417
- Wu, Y., Barquero, L. A., Pickren, S. E., Barber, A. T., & Cutting, L. E. (2020). The relationship between cognitive skills and reading comprehension of narrative and expository texts: A longitudinal study from Grade 1 to Grade 4. *Learning and Individual Differences*, 80, 101848. doi:10.1016/j.lindif.2020.101848
- Yildirim, K., Yildiz, M., & Ateş, S. (2011). Is vocabulary a strong variable predicting reading comprehension and does the

- prediction degree of vocabulary vary according to text types. *Educational Sciences: Theory and Practice*, 11(3), 1541-1547.
- Yim, D., & Han, J. (2019). Phonological loops, visuospatial sketchpad, episodic buffers, and inhibition: The relationship with grammar skills in children with a language delay. *Korean Journal of Special Education*, 54(2), 183-204. doi:10.15861/kjse.2019.54.2.183
- Yoon, H., Pae, S. K., & Chung, B. (2018). The role of vocabulary breadth and depth on reading comprehension in grade 1-4. *Communication Sciences & Disorders*, 23(2), 519-527. doi:10.12963/csd.18512
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2), 162-185. doi:10.1037/0033-2909.123.2.162

Appendix 1. Samples of literal and inferential comprehension questions

지문 종류	질문 유형	질문
문학	사실적 이해	• 람보는 무슨 동물인가요?(1점)
		• 람보는 자신의 어떤 점을 부끄러워 했나요?(1점)
		• 람보 엄마는 람보에게 뭐라고 위로해 주었나요?(2점)
		• 새들의 동네에 가기 전에, 람보는 무엇을 원했나요?(2점)
		• 이유는 무엇인가요?(2점)
	추론적 이해	• 람보는 엄마와 대화한 후에도 '외딴 섬에 떨어져 있는 것 같은 기분'이 들었다고 합니다. 어떤 기분인 것일까요?(2점)
		• 독수리 아저씨는 왜 람보를 새들의 동네로 데리고 갔을까요?(2점)
		• 독수리 아저씨의 말을 듣고 하늘을 날았을 때 람보는 왜 미소를 지었을까요?(2점)
		• 독수리 아저씨 등에 타고 집으로 돌아가는 길에 람보는 어떤 생각을 했을까요?(2점)
		• 람보는 새들의 동네에 다녀오면서 어떤 점을 깨닫게 되었을까요?(2점)
비문학	사실적 이해	• 세금은 누가 누구에게 내는 것인가요?(1점)
		• 세금은 그 나라에 사는 사람들 모두가 국가에 내야 하는 의무예요. 그런데 버는 돈이 너무 적은 사람은 어떻게 하나요?(1점)
		• 이 이야기에서는 국민들이 낸 세금으로 무엇을 한다고 했나요? 2가지 이상 말해 주세요(2점).
		• 만약 국민들이 세금을 내지 않는다면 지금과 다르게 어떤 일이 벌어질까요?(2점)
		• 어린이는 돈을 벌지 않지만, 사실은 세금을 내고 있는 것이나 다름없어요. 어떤 방식으로 세금을 내고 있는 것일까요?(2점)
	추론적 이해	• 세금은 학교, 도서관, 교과서, 횡단보도, 도로, 경찰서, 소방서 등에 사용됩니다. 이러한 것들의 공통적인 특징은 무엇인가요?(2점)
		• 만약 교과서를 세금으로 운영하지 않는다면, 어떤 일이 생길까요?(2점)
		• 수염세를 걷기 시작한 이후, 러시아 남자들은 어떻게 행동했을까요? 왜 그렇게 생각하나요?(2점)
		• 정체세를 거두기 시작한 이후 런던의 도로 정체가 줄어든 이유는 무엇일까요?(2점)
		• 런던의 '정체세'는 전기 자동차나 하이브리드 자동차에는 부과하지 않는다고 하였습니다. 그럼 전기 자동차나 하이브리드 자동차를 타는 사람은 많아질까요 아니면 적어질까요? 왜 그렇게 생각하나요?(2점)

텍스트 제시 양식, 지문 종류 및 질문 유형에 따른 초등학생 어휘발달지연 아동과 일반 아동의 판별력 분석

김신영^{1,2}, 임동선^{3*}

¹ 이화여자대학교 일반대학원 언어병리학과 강사

² (주)임동선아이세이언어연구소 언어재활사

³ 이화여자대학교 일반대학원 언어병리학과 교수

목적: 본 연구는 초등학교 1~6학년 아동들을 대상으로 텍스트 이해 능력을 검토하고, 텍스트 과제 조건 중 일반 아동과 어휘발달지연 아동을 가장 효과적으로 판별하는 조건이 무엇인지 검토하고자 하였다.

방법: 본 연구에는 초등학생 355명이 참여하였으며, 비언어성 지능검사 점수 결측 사례를 제외한 312명(일반 아동 285명, 어휘발달지연 아동 27명)을 최종 분석 대상으로 하였다. 연구 대상 아동들의 텍스트 이해 능력을 제시 양식(듣기, 듣기+읽기, 읽기), 지문 종류(문학, 비문학), 질문 유형(사실적 이해, 추론적 이해)의 과제 조건으로 평가하였으며, 이로부터 도출된 7개의 텍스트 과제 조건을 분석에 활용하였다.

결과: 연구 대상 집단을 유의하게 판별하는 텍스트 과제 조건이 무엇인지 검토하기 위하여 위계적 로지스틱 회귀분석을 실시하였으며, 그 결과 7개의 텍스트 과제 조건이 모두 집단 판별 모형의 설명력을 유의하게 증가시키는 것으로 나타났다($p < .05$). 추가적으로 ROC 및 AUC 분석 결과 민감도 측면에서는 '사실적 이해' 모형이(민감도=.704), 특이도 측면에서는 '읽기' 조건이(특이도=.916), 그리고 민감도와 특이도의 균형 측면에서는 '비문학 지문'이(AUC=.758, 민감도=.630, 특이도=.888) 가장 안정적인 분류 성능을 보이는 것으로 나타났다.

결론: 본 연구의 결과를 통해 각 텍스트 과제 조건이 독립적으로 집단 간 수행력 차이를 반영하는 의미 있는 지표임을 확인하였다. 구체적으로 사실적 이해 과제는 위험군 선별에, 읽기 조건은 일반 아동의 오분류 최소화, 비문학 지문은 민감도와 특이도의 균형을 고려한 평가에 유용한 것으로 나타났다. 이러한 결과를 고려할 때, 임상 및 교육 현장에서 텍스트 이해 평가 및 중재 시 단일 과제 유형에 의존하기보다, 평가 목적에 따라 전략적인 과제 조합을 구성할 필요가 있다.

검색어: 텍스트 이해, 어휘발달지연 아동, 판별력, 로지스틱 회귀분석, ROC 분석

교신저자: 임동선(이화여자대학교)

전자메일: sunyim@ewha.ac.kr

게재신청일: 2026. 05. 31

수정제출일: 2026. 07. 13

게재확정일: 2026. 07. 31

이 논문은 김신영(2026)의 박사학위 논문을 수정·보완하여 작성한 것임.

이 연구는 2021년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2021S1A3A2A01096102).

ORCID

김신영

<https://orcid.org/0000-0002-2294-2475>

임동선

<https://orcid.org/0000-0001-8254-9504>

참고 문헌

- 김영태, 홍경훈, 김경희, 장혜성, 이주연 (2009). **수용·표현 어휘력 검사**. 서울: 서울장애인종합복지관.
- 문수백 (2020). **한국판 카우프만 간면지능검사2(KBIT2)**. 서울: 인싸이트.
- 배미연 (2025). **독서 매체 유형 및 읽기 분량이 텍스트 이해에 미치는 영향: 초등학교 5학년을 중심으로**. 서울교육대학교 교육전문대학원 석사학위 논문.
- 윤효진, 배소영, 정부자 (2018). 초등 1-4학년 아동의 읽기이해능력에서 어휘지식의 양과 깊이의 역할. *Communication Sciences & Disorders*, 23(2), 519-527.
- 임동선, 한지운 (2019). 언어발달지체 아동의 음운루프, 시·공간 잡기장, 일화적 완충기, 억제기능과 문법 능력 간의 관계. *특수교육학연구*, 54(2), 183-204.
- 정하은, 송혜선, 조예림, 오유림, 주혜진, 임동선 (2023a). 이야기 과제 제시 조건에 따른 지문 유형과 질문 유형별 학령기 아동의 이야기 이해력 비교. *언어치료연구*, 32(2), 49-60.
- 정하은, 정은애, 남서현, 김정원, 윤서린, 김아영, 임동선 (2023b). 텍스트 유형 및 이야기 양식에 따른 초등학교 1-6학년의 이야기 이해 능력. *아동학회지*, 44(4), 389-400.