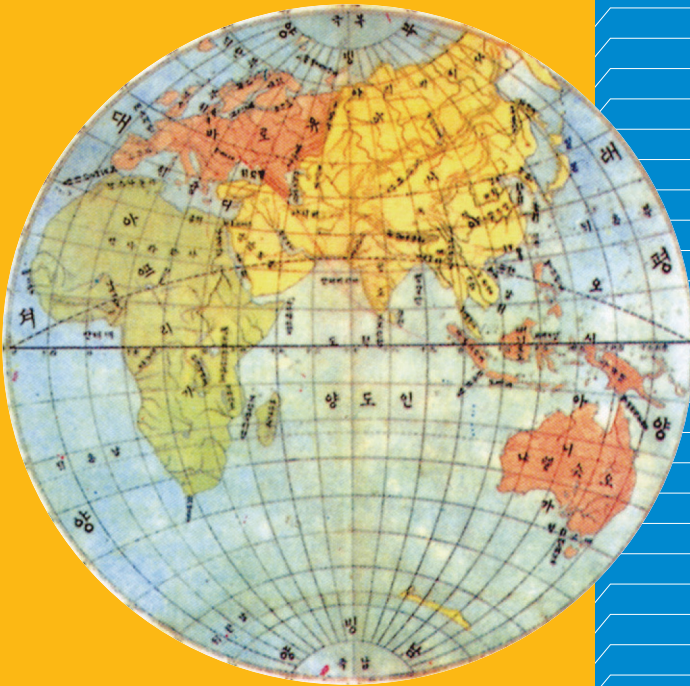


이중언어학

제104호



二重言語學會
2026. 6.

한국어-영어 이중언어 아동 어휘 검사 도구 VOKEB의 타당성 검토*

—한국어-영어 이중언어 아동을 위한 어휘력 평가 도구로서의 가능성—

김 아 영 · 김 민 지 · 박 민 수 · 임 동 선**

Abstract

Kim, Ayoung · Kim, Minji · Park, Minsu · Yim, Dongsun. 2026. 6. 30. **Validity Examination of VOKEB: A Vocabulary Assessment Tool for Korean-English Bilingual Children.** *Bilingual Research* 104, 1-18. This study examined the validity of VOKEB (Vocabulary of Korean-English Bilinguals) and the effects of composite scoring in bilingual vocabulary assessment. A total of 103 children aged 4 - 10 years (21 monolingual, 82 bilingual) were assessed using VOKEB, REVT, PPVT-4, and EOWPVT. VOKEB scores showed significant positive correlations with existing measures, supporting its validity. While bilingual children performed lower on raw scores, group differences diminished or reversed in receptive vocabulary when composite scoring was applied. The composite advantage ratio for VOKEB receptive vocabulary was significantly lower than that of other tests. These findings suggest that composite scoring provides a more equitable assessment of bilingual vocabulary and that VOKEB is a valid tool for measuring core vocabulary in bilingual children. (Ewha Womans University)

[Key words] bilingual children(이중언어 아동), vocabulary assessment(어휘 평가), composite scoring(합산 점수), VOKEB validity(VOKEB 타당도)

* 본 연구는 2025년 파라다이스재단의 지원을 받아 수행된 연구임

** 김아영: 제1저자, 김민지: 공동저자, 박민수: 공동저자, 임동선: 교신저자

1. 서론

다문화 사회로의 이행이 가속화됨에 따라, 이중언어 환경에서 성장하는 아동의 수는 전 세계적으로 빠르게 증가하고 있다. 특히 한국어와 영어를 병행하여 사용하는 이중언어 아동의 규모는 국내외를 막론하고 지속적으로 확대되고 있다. 2025년 재외동포청 자료에 따르면, 영어권 국가에 거주하는 한국계 동포는 미국 2,557,047명, 캐나다 263,153명, 영국 47,006명, 호주 170,215명, 뉴질랜드 39,155명 등 총 300만 명 이상에 이른다. 이들 가운데 상당수는 아동 및 청소년으로, 한글학교 및 한글교육원 재학생 수를 포함할 경우 한국어와 영어를 동시에 사용하는 아동·청소년 인구는 최소 341,850명 이상으로 추정된다(재외동포청, 2025). 더불어 국내에서도 제주도 국제학교, 유치부 영어학원, 영어 놀이학교 등에서 한국어와 영어를 병행 사용하는 아동의 수가 급증하고 있어, 한국어-영어 이중언어 아동의 규모는 향후 더욱 증가할 것으로 예상된다. 언어장애의 출현율이 3-9세 아동 기준 약 10% 내외라는 점(Choi et al., 2011; Tomblin et al., 1997)을 고려할 때, 약 35,000명 이상의 한국어-영어 이중언어 아동이 언어장애를 경험하고 있을 가능성이 있다. 이는 이중언어 아동을 대상으로 한 정확하고 타당한 언어 평가 도구의 필요성이 점차 커지고 있음을 시사한다.

이중언어 아동의 언어 습득 과정은 단일언어 아동과 비교할 때 구조적, 환경적, 인지적 측면에서 다른 특성을 지닌다. 이들은 언어 간 전이(cross-linguistic transfer), 언어 간 비대칭성(asymmetry), 개념 지식의 분산적 표현(distributed vocabulary knowledge)과 같은 고유한 언어 특성을 보일 수 있다. 단일언어 아동이 하나의 언어 체계 내에서 지속적이고 집중적인 언어 노출을 통해 어휘를 습득하는 것과 달리, 이중언어 아동은 두 언어에 걸쳐 언어 노출과 상호작용의 기회를 분산적으로 경험한다. 이러한 환경은 각 언어에서의 어휘 습득 속도를 상대적으로 늦출 수 있

으나, 두 언어를 모두 활용하여 보다 넓은 범주의 개념을 습득하게 된다(Pearson et al., 1993). 이중언어 아동은 특정 개념을 한 언어로는 알고 있으나 다른 언어로는 알지 못하는 양상을 보일 수 있으며, 이는 두 언어에 대한 노출 및 사용 빈도와 환경적 맥락의 차이에 따라 결정된다(Kan & Kohnert, 2005; Kohnert, 2010; Bedore et al., 2011). 이러한 개념 지식의 분산은 단일언어 기반 평가 도구를 사용할 경우 아동의 실제 언어 능력이 과소평가될 가능성을 내포한다. 특히 진단적 판단이 요구되는 임상 상황에서는 이러한 평가상의 왜곡이 언어장애의 오진으로 이어질 위험이 크다(Peña et al., 2001).

이러한 한계를 보완하기 위한 대안으로 제안된 방식이 합산 점수(composite scoring)이다. 합산 점수는 두 언어 중 어느 하나라도 특정 개념을 정확히 표현할 수 있을 경우 이를 정답으로 인정하는 채점 방식으로, 언어 형태가 아닌 개념 지식 자체를 평가의 핵심으로 삼는다(Pearson et al., 1997). 이 방식은 이중언어 아동의 어휘 능력을 보다 공정하게 반영할 수 있다는 점에서 중요한 의의를 지닌다(Bedore et al., 2005). 예컨대, 단일언어 점수만을 기준으로 할 경우 ‘tree’라는 단어를 영어로 표현하지 못하는 한국어-영어 이중언어 아동이 언어 지체로 오인될 수 있으나, 해당 아동이 한국어로 ‘나무’라는 개념을 정확히 알고 있다면 이는 전혀 다른 진단 결과를 초래할 수 있다. 합산 점수방식의 도입은 단일언어 평가에서 포착되지 않는 개념적 어휘 지식을 반영함으로써 평가의 타당도와 진단 정확도를 향상시킨다.

더 나아가 합산 점수는 임상 현장에서의 실천 가능성 측면에서도 중요한 의미를 갖는다. 합산 점수를 적용한 체계적인 평가 절차와 의사결정 지침은 이중언어 아동의 언어장애 여부를 보다 정확히 판별하는 데 기여할 수 있다(Peña, Bedore, & Kester, 2016). 실제로 개념 단위 채점, 전문가 합의 절차 등을 포함한 합산 점수 적용 연구에서는 언어장애 진단의 민감도와 신뢰도가 동시에 향상됨이 보고되었다(McClain et al., 2021).

또한 부모 및 교사 보고, 언어 노출 환경, 사회문화적 맥락을 함께 고려하는 다차원적 평가 접근이 권고되며(Castilla-Earls et al., 2020), 두 언어 수행을 동시에 고려하는 이차원적 임상 의사결정 모형(Bedore et al., 2011)과 best-language 또는 평균 점수 기반 합산 점수 접근(Peña et al., 2018)은 이러한 평가의 대표적인 예로 제시된다.

그럼에도 불구하고, 현재 이중언어 아동을 대상으로 한 합산 점수 기반 어휘 평가는 여전히 여러 한계를 지닌다. 한국어-영어 이중언어 아동에게 주로 사용되는 REVT(Receptive & Expressive Vocabulary Test; Kim et al., 2009), PPVT-4(Dunn & Dunn, 2007), EOWPVT-4(Martin & Brownell, 2010)는 모두 단일언어 화자를 대상으로 개발된 검사로, 이중언어 환경을 반영한 문항 구성이나 규준이 포함되어 있지 않다. 그 결과, 합산 점수 산출을 위해 복잡하고 장시간의 평가 절차가 요구되며, 이는 아동과 검사자 모두에게 높은 부담을 초래한다. 또한 이중언어 아동 집단의 수행 데이터를 기반으로 한 표준화 규준이 부재하다는 점에서 평가의 타당성 역시 제한적이다.

이러한 문제에도 불구하고, 현재까지 한국어-영어 이중언어 아동을 대상으로 한 규준 기반 어휘 검사 도구나, 문항의 난이도 및 변별도를 고려한 필수 문항 중심의 평가 도구 개발은 충분히 이루어지지 않았다. 이에 본 연구에서는 한국어-영어 이중언어 아동의 어휘 능력을 평가하기 위한 새로운 검사 도구인 VOKEB(Vocabulary of Korean-English Bilinguals)를 제안한다. VOKEB는 문화적 중립성을 고려한 시각 자극과 어휘 문항을 사용하며, 합산 점수 방식을 적용하여 두 언어 중 어느 하나로라도 개념을 정확히 표현할 수 있을 경우 이를 정답으로 인정하는 구조를 채택하였다. 이러한 접근은 이중언어 아동의 다양한 언어 경험을 반영하고 평가의 공정성과 정확성을 동시에 확보할 수 있다는 점에서 학술적·임상적 의의를 지닌다.

본 연구는 VOKEB의 타당성을 검증하기 위해 기존 표준화 단일언어

검사 도구인 REVT, PPVT-4, EOWPVT 점수와의 상관을 분석하고, 검사별 합산 점수로 인한 이득 점수 비율(composite advantage ratio)을 비교하고자 한다. 이를 통해 단일언어 기반 평가와 합산 점수 기반 평가 간의 성능 차이를 정량적으로 검증하고, 다문화·다언어 환경에서의 어휘 평가 실천을 보다 공정하고 효과적으로 개선할 수 있는 근거를 제시하고자 한다. 본 연구의 연구 질문은 다음과 같다.

1. 단일언어 아동과 이중언어 아동 간 어휘 점수에는 유의한 차이가 있는가?
2. VOKEB과 기존 어휘 검사 도구(REVT, PPVT-4, EOWPVT) 간에는 유의한 상관이 존재하는가?
3. 검사 도구별 합산 점수 이득 비율에는 유의한 차이가 존재하는가?

2. 연구 방법

2.1. 연구 대상

만 4세에서 10세 사이의 아동 총 103명이 본 연구에 참여하였다. 이들은 한국어 단일언어 아동 21명과 한국어-영어 이중언어 아동 82명으로 구성되었다. 한국어-영어 이중언어 아동은 부모의 보고에 따라 가정에서는 한국어를 모국어로 사용하고, 교육기관 및 사회적 환경에서는 영어를 사용하는 아동으로 정의하였다. 모든 아동은 인지 및 청력에 문제가 없었으며, 한국 카우프만 간편지능 검사-2(Korean Kaufman Brief Intelligence Test-Second Edition; KBIT-2, Moon, 2020)를 통해 비언어성 지능 지수가 85점(-1SD) 이상임을 확인하였다. 또한, 한국어 단일언어 아동의 월령($M = 70.43$, $SD = 22.64$)과 한국어-영어 이중언어 아동의 월령($M = 75.94$, $SD = 18.39$) 간에는 독립표본 t 검정 결과 유의한

차이가 없었다.

2.2. 연구 도구

아동의 전반적인 어휘 능력을 평가하기 위해, 현재 개발 중인 어휘 검사 도구인 VOKEB과 표준화된 한국어 어휘 검사인 REVT(Receptive and Expressive Vocabulary Test; 김영태 외, 2006)를 실시하였다. VOKEB은 만 3세에서 10세 아동을 대상으로 수용 및 표현 어휘 능력을 평가하기 위해 개발 중인 검사로, 표현 어휘 문항 170개(파일럿 버전 155개)와 수용 어휘 문항 179개(파일럿 버전 163개)로 구성되어 있다. 본 연구는 파일럿 버전을 사용하여 진행하였다.

한국어-영어 이중언어 아동의 어휘 능력을 보다 정밀하게 평가하기 위해, 영어 수용 어휘 검사인 PPVT-4(Peabody Picture Vocabulary Test-Fourth Edition; Dunn & Dunn, 2007)와 영어 표현 어휘 검사인 EOWPVT (Expressive One-Word Picture Vocabulary Test; Martin & Brownell, 2011)를 추가로 실시하였다.

이중언어 아동의 경우 합산 점수(composite score) 산출을 위해, 한국어 어휘 검사인 REVT(김영태 외, 2006)의 수용 어휘 문항 185개와 영어 어휘 검사인 PPVT-4(Dunn & Dunn, 2007)의 수용 어휘 문항 228개를 대상으로 번역 절차를 실시하였다. 번역은 다음과 같은 3단계 절차를 통해 이루어졌다.

1단계에서는 옥스포드 영한사전을 기준으로 각 어휘의 사전적 의미를 최대 세 번째 의미까지 번역하였다. 해당 어휘가 수록되어 있지 않은 경우에는 동아 영한사전과 YBM 영한사전을 추가로 참고하였다. REVT 수용 어휘 문항 역시 동일한 절차에 따라 영어로 번역되었다.

2단계에서는 번역된 어휘의 연령 적절성을 검증하기 위해 한국어가 모국어인 유치원 교사 1인과 영어가 모국어인 중학교 교사 1인에게 평가

를 의뢰하였다. 각 어휘가 목표 연령에 적합한지를 수치화하여 평가하였으며, 이를 통해 연령 적절성을 확인하였다.

3단계에서는 번역 어휘의 그림-어휘 일치성을 검증하기 위해 각 언어가 모국어인 성인을 대상으로 해당 어휘를 모국어로 번역하도록 하여, 번역 결과가 제시된 어휘와 일치하는지를 분석하였다. 최종 어휘 선정은 한국과 미국 공인 언어재활사이자 한국어와 영어가 모두 모국어인 언어병리학 교수 1인, 1급 언어재활사이며 두 언어에 능통한 박사과정생 2인과 석사과정생 2인으로 구성된 전문가 집단이 수행하였다. 전문가들은 각 문항에서 제시된 번역 중 의미적·문화적 적절성이 가장 높은 어휘를 선정하였다.

선정된 어휘는 한국어가 모국어인 교사 1인과 영어가 모국어인 교사 1인을 통해 연령 및 문화적 적절성에 대한 추가 검토를 거쳐 최종 확정되었다.

2.3. 연구 절차

모든 검사는 Zoom을 통해 비대면으로 실시하였다. 이중언어 아동의 합산 점수 산출을 위한 검사 절차는 검사 도구의 언어 특성에 따라 다음과 같이 진행하였다. 표현어휘의 경우, REVT와 VOKEB는 한국어로 먼저 실시한 후 정반응·오반응 문항을 혼합하여 동일한 개념을 영어로 재제시하였으며, EOWPVT는 영어로 먼저 실시한 후 정반응·오반응 문항을 혼합하여 한국어로 재제시하였다. 수용어휘의 경우에도 동일한 원칙을 적용하여, REVT와 VOKEB는 한국어로 먼저 실시한 후 정반응·오반응 문항을 혼합하여 영어로 재제시하였고, PPVT-4는 영어로 먼저 실시한 후 정반응·오반응 문항을 혼합하여 한국어로 재제시하였다. 특히 수용어휘 검사는 4지선다 형식으로 구성되어 있어, 재제시 시 아동이 이미 오반응한 보기를 제외하고 제시함으로써 단순한 확률적 이점으로 인한

점수 상승 가능성을 통제하였다. 두 언어 중 어느 하나에서라도 정반응한 문항은 정답으로 인정하여 합산 점수를 산출하였다.

3. 연구 결과

3.1. 기술 통계

본 연구에서는 단일언어 아동과 이중언어 아동의 어휘 수행 수준을 비교하기 위해 각 검사별 기술통계를 산출하였다.

단일언어 아동 집단($n=21$)의 경우, VOKEB 표현 점수는 평균 122.76($SD=18.11$), 수용 점수는 평균 127.05($SD=22.97$)로 나타났다. REVT에서는 표현 점수 평균이 71.29($SD=14.59$), 수용 점수 평균이 69.62($SD=22.79$)로 나타났다.

이중언어 아동 집단($n=82$)의 원점수의 경우, VOKEB 표현 점수는 평균 111.90($SD=28.51$), 수용 점수는 평균 130.78($SD=19.26$)로 나타났다. REVT에서는 표현 점수 평균이 68.52($SD=28.84$), 수용 점수 평균이 67.04($SD=27.09$)로 나타났으며 EOWPVT 평균은 70.44($SD=26.65$), PPVT-4 평균은 89.82($SD=36.93$)로 확인되었다.

한편, 이중언어 아동의 합산 점수를 살펴보면 VOKEB 표현 점수는 평균 129.57($SD=17.11$), 수용 점수는 평균 146.77($SD=15.18$)로 나타나 이중언어 점수보다 높은 수행을 보였다. REVT에서는 표현 점수 평균이 84.07($SD=25.42$), 수용 점수 평균이 95.16($SD=33.78$)로 나타났으며, EOWPVT 평균은 85.24($SD=21.15$), PPVT-4 평균은 122.02($SD=39.97$)로 나타났다.

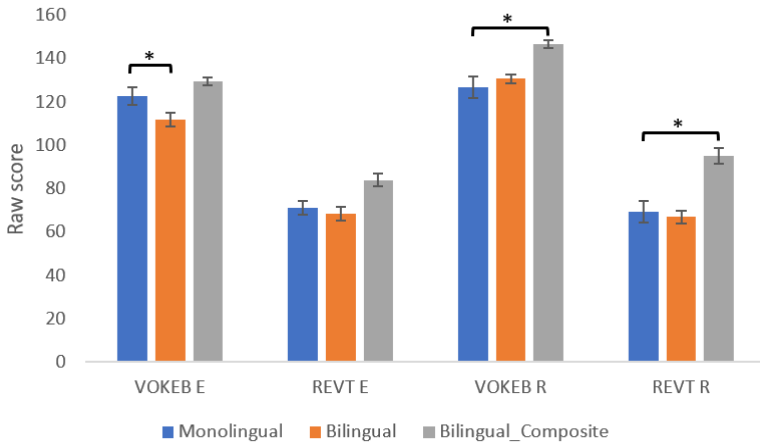
전반적으로 이중언어 아동은 원점수에서는 상대적으로 낮은 수행을 보였으나, 합산 점수를 적용할 경우 전반적인 어휘 수행 수준이 향상되는 경향을 확인할 수 있었다.

<표 1> 기술통계

	VOKEB_E	VOKEB_R	REVT_E	REVT_R	EOWPVT	PPVT-4
단일언어 (n=21)	122.76 (18.19)	127.05 (22.97)	71.29 (14.59)	69.62 (22.79)	-	-
이중언어 원점수 (n=82)	111.9 (28.51)	130.78 (19.21)	68.52 (28.84)	67.04 (27.09)	70.44 (26.65)	89.82 (36.93)
이중언어 합산 점수 (n=82)	129.57 (17.153)	146.77 (15.11)	84.07 (25.42)	95.16 (33.7)	85.24 (21.15)	122.02 (39.97)

3.2. 단일언어 아동 집단과 이중언어 아동 집단 간 어휘 점수 차이

독립표본 t 검정을 사용하여 단일언어 아동 집단과 이중언어 아동 집단 간 어휘 점수에 차이가 유의한지 살펴보았다. 먼저 표현어휘 원점수의 경우 REVT에서는 두 집단 간 차이가 유의하지 않았으나 VOKEB에서는 두 집단 간 차이가 유의하였다($t(48.359)=2.143$ $p=.037$). 즉, 단일언어 아동 집단 ($M=122.76$, $SD=18.19$)이 이중언어 아동 집단($M=111.9$, $SD=28.51$)에 비해 표현 어휘의 원점수가 유의하게 높은 것으로 나타났다. 그러나 합산 점수로 비교하였을 때 단일언어 집단과 이중언어 집단 간 차이가 유의하지 않았다($p>.05$). 수용어휘의 원점수로 비교하였을 때는 단일언어 아동 집단 (VOKEB: $M=127.05$, $SD=22.97$, REVT: $M=69.62$, $SD=22.79$)과 이중언어 아동 집단(VOKEB: $M=130.78$, $SD=19.21$, REVT: $M=67.04$, $SD=27.09$) 간 차이가 유의하지 않았으나($p>.05$), 합산점수로 비교하였을 때는 이중언어 아동(VOKEB: $M=146.77$, $SD=15.11$, REVT: $M=95.16$, $SD=33.7$)과 단일언어 아동 간 유의한 차이가 나타났다(VOKEB: $t(101)=-4.755$, $p<.001$, REVT: $t(45.185)=-4.112$, $p<.001$). 이는 합산점수로 수용 어휘를 살펴보았을 때 이중언어 아동 집단이 단일언어 아동 집단보다 더 많은 단어를 맞혔다고 해석된다.



<그림 1> 어휘 검사에 따른 집단 별 어휘점수

3.3. VOKEB과 기존 어휘 검사들간의 상관

VOKEB과 기존 검사들 간의 상관 관계를 알아보기 위해 Pearson 상관 분석을 실시하였다. 결과는 아래 표와 같다. 먼저 VOKEB 표현 어휘의 원 점수는 REVT 표현 어휘의 원점수($r=.842, p<.001$), 합산 점수($r=.511, p<.001$), 수용어휘의 원점수($r=.697, p<.001$), 합산 점수($r=.329, p<.001$)와 유의한 정적 상관관계가 있었다. 또한 EOWPVT의 합산점수($r=.294, p=.007$)와 PPVT-4의 합산점수($r=.297, p=.007$)와도 유의한 상관이 존재하였다. VOKEB 표현 어휘의 합산 점수는 REVT 표현 어휘의 원점수($r=.598, p<.001$), 합산 점수($r=.761, p<.001$), 수용 어휘의 원점수($r=.689, p<.001$), 합산점수($r=.727, p<.001$), EOWPVT의 원점수($r=.657, p<.001$), 합산점수($r=.753, p<.001$), PPVT-4의 원점수($r=.674, p<.001$), 합산점수($r=.778, p<.001$)와 모두 유의한 정적 상관관계가 있었다. VOKEB 수용 어휘의 원점수는 REVT 표현 어휘의 원점수($r=.703, p<.001$), 합산 점수($r=.691, p<.001$), 수용 어휘의 원점수($r=.803, p<.001$), 합산점수($r=.660,$

$p<.001$), EOWPVT의 원점수($r=.349$, $p<.001$), 합산점수($r=.546$, $p<.001$), PPVT-4의 원점수($r=.500$, $p<.001$), 합산점수($r=.651$, $p<.001$)와 모두 유의한 정적 상관관계가 있었다. 마지막으로 VOKEB 수용 어휘의 합산 점수는 REVT 표현 어휘의 원점수($r=.390$, $p<.001$) 및 합산점수($r=.644$, $p<.001$), 수용 어휘의 원점수($r=.567$, $p<.001$) 및 합산점수($r=.742$, $p<.001$), EOWPVT의 원점수($r=.623$, $p<.001$) 및 합산점수($r=.678$, $p<.001$), PPVT-4의 원점수($r=.711$, $p<.001$) 및 합산점수($r=.767$, $p<.001$)와 유의한 정적 상관관계가 있었다.

<표 2> 상관분석표

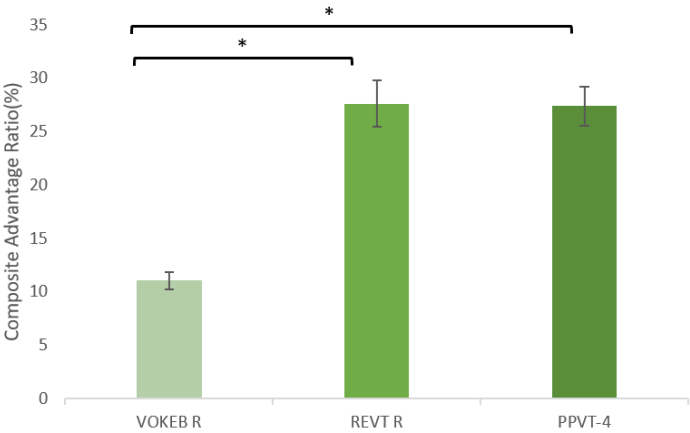
	REVT_E 원점수	REVT_E 합산점수	REVT_R 원점수	REVT_R 합산점수	EOWPVT 원점수	EOWPVT 합산점수	PPVT-4 원점수	PPVT-4 합산점수
VOKEB_E 원점수	.842***	.511***	.687***	.329***	.056	.294**	.128	.297**
VOKEB_E 합산점수	.598***	.761***	.689***	.727***	.657***	.753***	.674***	.778***
VOKEB_R 원점수	.703***	.691***	.803***	.660***	.349***	.546***	.500***	.651***
VOKEB_R 합산점수	.390***	.644***	.567***	.742***	.623***	.678***	.711***	.767***

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

3.4. 검사별 이득점수비율 간 차이

먼저 VOKEB 표현어휘($M=13.88$, $SD=17.25$) REVT 표현어휘($M=18.69$, $SD=21.86$), EOWPVT($M=18.92$, $SD=19.15$)의 이득점수비율 간 차이가 있는지 알아보기 위해 반복측정 분산분석을 시행한 결과, 세 비율 간 유의한 차이가 나타나지 않았다($p>.05$). 그러나 VOKEB 수용어휘($M=11.05$, $SD=7.6$), REVT 수용어휘($M=27.65$, $SD=19.46$), PPVT-4($M=27.40$, $SD=16.90$)의 이득점수비율 간 차이가 있는지 살펴보았을 때, 세 비율 간

유의한 차이가 있는 것으로 나타났다($F(1.335, 108.13)=8.276, p<.001$). Bonferroni 사후검정을 실시한 결과, VOKEB 수용어휘에서의 이득점수 비율이 REVT($p<.001$)와 PPVT-4($p<.001$)에서의 이득점수비율보다 유의하게 낮아 나타난 결과로 해석되었다.



<그림 2> 어휘 검사 별 이득점수비율 간 차이

4. 결론

본 연구는 한국어-영어 이중언어 아동을 대상으로 개발 중인 어휘 검사 도구 VOKEB의 타당성을 검증하고, 합산 점수(composite scoring) 적용 시 단일언어 아동과 이중언어 아동 간 어휘 수행의 차이를 살펴보고자 하였다. 연구 결과는 크게 세 가지 측면에서 논의될 수 있다.

첫째, VOKEB 점수는 기존의 표준화된 어휘 검사(REVT, PPVT-4, EOWPVT) 점수와 전반적으로 높은 정적 상관을 보였다. 특히 VOKEB의 합산 점수는 한국어 및 영어 어휘 검사들의 합산 점수와 일관되게 높은 상관을 나타내어, 본 도구가 기존 검사들과 동일한 어휘 능력 구성 개

념을 측정하고 있음을 시사한다.

어휘 검사 개발 연구에서 기존 표준화 검사와의 상관 분석은 수렴 타당성을 검증하는 대표적인 방법으로 활용되어 왔다(Dunn & Dunn, 2007; Ingvalson et al., 2023). 본 연구에서 확인된 상관 계수는 기존 어휘 검사 간 상관 수준과 유사하거나 이에 준하는 범위로, VOKEB이 임의의 과제를 측정하는 것이 아니라 기존 검사들이 평가해 온 어휘 능력과 본질적으로 연관된 구인을 측정하고 있음을 뒷받침한다.

둘째, 원점수 기준에서는 이중언어 아동이 단일언어 아동보다 표현어휘에서 유의하게 낮은 수행을 보였으나, 합산 점수를 적용한 이후에는 두 집단 간 차이가 유의하지 않게 나타났다. 이러한 결과는 표현어휘뿐 아니라 수용어휘에서도 유사한 경향을 보였으며, 특히 수용어휘의 경우 합산 점수 적용 시 이중언어 아동의 평균 점수가 단일언어 아동을 상회하였다.

이는 이중언어 아동의 어휘 지식이 두 언어에 분산되어 나타난다는 선행 연구의 주장과 일치한다(Pearson et al., 1993; Bedore et al., 2005). 단일언어 기반 평가에서는 특정 개념을 한 언어로만 알고 있는 경우 정반응으로 인정되지 않아 이중언어 아동의 실제 어휘 능력이 과소평가될 수 있다. 반면, 합산 점수 방식은 언어 형태가 아닌 개념 단위의 지식을 평가함으로써 이중언어 아동의 언어 경험을 보다 공정하게 반영한다.

특히 수용어휘 영역에서 합산 점수 적용 시 집단 간 차이가 사라지거나 역전되는 결과는, 수용어휘가 표현어휘에 비해 언어적 부담이 적고 개념 지식 자체를 보다 직접적으로 반영하기 때문일 가능성이 있다(Gross et al., 2014). 이러한 결과는 합산 점수 방식이 이중언어 아동의 수용어휘 평가에서 특히 유용한 접근임을 시사한다.

셋째, 검사별 이득점수비율을 비교한 결과 VOKEB 수용어휘의 이득점수비율은 REVT와 PPVT-4에 비해 유의하게 낮은 것으로 나타났다. 이득점수비율은 합산 점수 적용을 통해 추가적으로 정반응으로 인정된 문항의 비율을 의미하며, 일반적으로 이 비율이 높을수록 한 언어에서만

알고 있는 단어가 많음을 시사한다.

VOKEB에서 이득점수비율이 상대적으로 낮게 나타난 결과는, 본 도구의 문항이 특정 언어 환경에 특화된 어휘보다는 언어 간 공통적으로 습득되는 핵심 어휘, 즉 아동의 일상적 언어 환경에서 높은 빈도로 사용되며 특정 문화권에 편향되지 않는 기본 개념 어휘를 중심으로 구성되어 있기 때문일 가능성을 시사한다. 선행 연구에 따르면 고빈도이자 개념적으로 보편적인 어휘는 단일언어 아동과 이중언어 아동 모두에게 비교적 안정적으로 습득되며, 이러한 어휘일수록 언어 간 분산 효과로 인한 이득 점수가 적게 나타난다(Sullivan et al., 2018).

특히 수용어휘 영역에서 이러한 경향이 두드러졌다는 점은, VOKEB이 아동의 기본적인 개념 어휘 지식을 평가하는 데 초점을 둔 도구임을 시사한다. 이는 단순히 두 언어의 어휘를 합산하여 점수를 극대화하는 검사라기보다는, 이중언어 아동의 핵심 어휘 능력을 보다 안정적으로 측정하려는 설계가 잘 이루어진 것으로 판단된다.

본 연구의 한계점으로는 합산 점수 산출을 위해 번역된 어휘 문항 간의 문화적인 편향을 완전히 통제하기 어렵다는 점을 들 수 있다. 비록 본 연구에서는 다단계 번역 절차와 전문가 합의를 통해 의미적, 문화적 적절성을 최대한 확보하고자 하였으나, 언어 및 문화적 맥락 차이로 인해 일부 어휘는 두 언어에서 동일한 수준이 아닐 가능성이 있다. 이러한 제한은 합산 점수 방식 전반에 내재된 구조적 한계로, 향후 개념 단위 문항 개발이나 문화 중립적 자극을 활용한 검사 설계를 통해 보완될 필요가 있다.

또 다른 한계로는 두 집단 간 표본 크기의 불균형을 들 수 있다. 본 연구에서 한국어 단일언어 아동은 21명, 한국어-영어 이중언어 아동은 82명으로 구성되어 두 집단 간 표본 수의 차이가 크며, 특히 단일언어 아동의 표본 수가 적어 결과의 일반화 가능성이 제한될 수 있다. 이러한 불균형은 집단 간 비교 시 통계적 검정력에 영향을 미칠 수 있으며, 결과 해석 시 이 점이 고려되어야 한다. 또한 본 연구에서는 언어 노출량, 사용

환경, 교육 환경 등의 환경 변인이 충분히 통제되지 않았다는 점도 한계로 지적될 수 있다. 단일언어 아동과 이중언어 아동 모두 이러한 환경 변인에 따라 어휘 발달의 개인차가 크게 나타날 수 있으므로, 향후 연구에서는 보다 균형 잡힌 표본 구성과 함께 환경 변인에 대한 체계적인 통제가 이루어질 필요가 있다.

그럼에도 불구하고, 본 연구는 합산 점수 방식을 적용한 VOKEB이 단일언어 기반 평가에서 이중언어 아동의 어휘 능력이 과소평가되는 문제를 완화하고, 기존 표준화 어휘 검사들과의 높은 상관을 통해 타당한 어휘 평가 도구로 기능할 수 있음을 보여주었다. 특히 수용어휘 영역에서 나타난 결과는 VOKEB이 이중언어 아동의 핵심 개념 어휘를 안정적으로 반영하는 도구로서, 다언어 환경에서의 공정한 어휘 평가를 위한 실질적인 대안을 제시함을 시사한다. 본 연구는 VOKEB의 파일럿 버전을 사용한 초기 타당성 검토 연구로서, 현재 기준 자료 확보를 위한 대규모 데이터 수집 및 아동 발화 자료 수집이 진행 중에 있다. 이를 통해 연령 대별 기준을 마련하고 문항별 난이도 및 변별도를 정교화하는 표준화 연구가 후속적으로 이루어질 예정이며, VOKEB이 임상 및 교육 현장에서 보다 공정하고 신뢰로운 이중언어 아동 어휘 평가 도구로 자리매김할 수 있기를 기대한다.

<참고문헌>

- 김영태, 홍경훈, 김경희, 장혜성, 이주연(2009). 『수용·표현어휘력검사』. 서울: 서울장애인종합복지관.
- 문수백(2020). 『한국판 카우프만 간편지능검사 2』. 서울: 학지사.
- 재외동포청(2025). 2025년 재외동포 현황. <재외동포청>. www.oka.go.kr
- Bedore, L. M., Pena, E. D., Garcia, M., & Cortez, C. (2005). Conceptual versus monolingual scoring. *Language, Speech, and Hearing Services in Schools*, 36(3), 188-200.

- Bedore, L. M., Pena, E. D., Joyner, D., & Macken, C. (2011). Parent and teacher rating of bilingual language proficiency and language development concerns. *International Journal of Bilingual Education and Bilingualism*, 14(5), 489-511.
- Castilla-Earls, A., Bedore, L., Rojas, R., Fabiano-Smith, L., Pruitt-Lord, S., Restrepo, M. A., & Peña, E. (2020). Beyond scores: Using converging evidence to determine speech and language services eligibility for dual language learners. *American journal of speech-language pathology*, 29(3), 1116-1132.
- CHOI, J. Y., KIM, C. A., SONG, I. J., LEE, K. W., GANG, M. J., JUNG, M. J., & SON, B. H. (2011). Clinical review of children diagnosed as specific language impairment. *Journal of the Korean Child Neurology Society*, 19(1), 8-17.
- Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test - Fourth Edition (PPVT-4)*. Pearson.
- Gross, M., Buac, M., & Kaushanskaya, M. (2014). Conceptual scoring of receptive and expressive vocabulary measures in simultaneous and sequential bilingual children. *American journal of speech-language pathology*, 23(4), 574-586.
- Ingvalson, E. M., Perry, L. K., VanDam, M., & Grieco-Calub, T. M. (2023). Comparing Scores on the Peabody Picture Vocabulary Test and Receptive One-Word Picture Vocabulary Test in Preschoolers With and Without Hearing Loss. *American journal of speech-language pathology*, 32(4), 1610-1619.
- Kan, P. F., & Kohnert, K. (2005). Preschoolers learning Hmong and English: Lexical-semantic skills in L1 and L2. *Journal of Speech, Language, and Hearing Research*, 48(2), 372-383.
- Kohnert, K. (2010). Bilingual children with primary language impairment: Issues, evidence and implications for clinical actions. *Journal of communication disorders*, 43(6), 456-473.
- Martin, N. A., & Brownell, R. (2011). *Expressive One-Word Picture Vocabulary Test, Fourth Edition (EOWPVT-4)*. Academic Therapy Publications.
- McClain, J. B., Oh, M. H., & Mancilla & Martinez, J. (2021). Questioning the monolingual norm with conceptually scored bilingual vocabulary assessments: Findings from a research-practice partnership. *TESOL Journal*, 12(3), e585.
- Pearson, B. Z., Fernández, S. C., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language learning*, 43(1), 93-120.
- Pearson, B. Z., Fernandez, S. C., Lewedeg, V., & Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied psycholinguistics*, 18(1), 41-58.

- Peña, E. D., Bedore, L. M., & Baron, A. (2017). Bilingualism and child language disorders. In *Handbook of child language disorders* (pp. 297-327). Psychology Press.
- Pena, E. D., Bedore, L. M., & Kester, E. S. (2016). Assessment of language impairment in bilingual children using semantic tasks: Two languages classify better than one. *International Journal of Language & Communication Disorders*, 51(2), 192-202.
- Sullivan, M. D., Poarch, G. J., & Bialystok, E. (2018). Why is Lexical Retrieval Slower for Bilinguals? Evidence from *Picture Naming. Bilingualism* (Cambridge, England), 21(3), 479-488.
- Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of speech, language, and hearing research*, 40(6), 1245-1260.
- Wood, C., Schatschneider, C. (2019). Item bias: Predictors of accuracy on peabody picture vocabulary test-fourth edition items for Spanish-English-speaking children. *Journal of Speech, Language and Hearing Research*, 62, 1392-1402.

김아영(Kim, Ayoung)(제1저자)

이화여자대학교 언어병리학과 석사과정생

(03760) 서울특별시 서대문구 이화여대길 52 교육관 A동

전화번호: 02-3277-6720

전자우편: zzzz2791@ewha.ac.kr

김민지(Kim, Minji)(공동저자)

이화여자대학교 언어병리학과 석사과정생

(03760) 서울특별시 서대문구 이화여대길 52 교육관 A동

전화번호: 02-3277-2120

전자우편: minji.k@ewha.ac.kr

박민수(Park, Minsu)(공동저자)

이화여자대학교 언어병리학과 석사과정생

(03760) 서울특별시 서대문구 이화여대길 52 교육관 A동

전화번호: 02-3277-2120

전자우편: scarlettee@ewha.ac.kr

18 이중언어학 제104호(2026)

임동선(Yim, Dongsun)(교신저자)

이화여자대학교 언어병리학과 교수

(03760) 서울특별시 서대문구 이화여대길 52 교육관 A동

전화번호: 02-3277-2120

전자우편: sunyim@ewha.ac.kr

접수일자: 2026년 4월 28일

심사(수정)일자: 2026년 5월 28일

게재확정: 2026년 6월 9일