JSLHD
2025. Vol. 34, No. 4. 043-051
ISSN 1226-587X / eISSN 2671-7158
Journal of Speech-Language & Hearing Disorders 언어치료연구

ORIGINAL ARTICLE
https://doi.org/10.15724/jslhd.2025.34.4.043
http://jslhd.org/

The Validity of Frequently Occurring Vocabulary for Toddlers Based on Generative AI

Si Hyun Kim[1*]

[1] *Dept. of Speech-Language Pathology, Graduate School of Health & Welfare, Konyang University, Professor*

Purpose : The purpose of this study was to examine the validity of a high-frequency vocabulary list for toddlers generated using generative artificial intelligence from multiple perspectives, and to explore the potential of artificial intelligence as a new approach in vocabulary development research.

Methods : A high-frequency vocabulary list for 18~36-month-old toddlers was generated using ChatGPT-4.0. Eleven experts participated in evaluating its validity in terms of developmental, functional, and frequency aspects. Vocabulary items were selected based on validity criteria, and the extraction ratio of the selected items was analyzed. In addition, the selected vocabulary was compared with the Korean MacArthur-Bates Communicative Development Inventories (K M-B CDI), a standardized vocabulary assessment tool, to analyze the overlap rate.

Results : First, the 486 vocabulary items generated by ChatGPT-4.0 generally met the validity criteria across developmental, functional, and frequency perspectives. Second, a total of 313 vocabulary items were identified as valid, corresponding to 64.40% of the entire list. Specifically, the selected items included the following words: 24 for people, 28 for body, 39 for animal, 42 for food, 53 for object, 48 for action, 37 for state, 15 for place, 10 for nature, 7 for color, 5 for number, and 5 for shape. Third, the overlap rate between the selected vocabulary and the K M-B CDI list was 77.00%.

Conclusions : These findings suggest that AI-based vocabulary lists can complement the scope of existing standardized assessment tools while reflecting the changing language environment. This study demonstrates the potential of AI in vocabulary development research and provides foundational data that can be applied as reference material for evaluation and intervention in early childhood education and speech-language rehabilitation.

Keywords : Generative AI, toddler, vocabulary, validity

Correspondence : Si Hyun Kim, PhD
E-mail : shkim1881@konyang.ac.kr

Received : September 11, 2025
Revision revised : September 24, 2025
Accepted : October 31, 2025

This paper was supported by the Konyang University Research Fund in the first half of 2024.

ORCID
Si Hyun Kim
https://orcid.org/0000-0002-2476-4529

Ⅰ. 서론

영유아기의 어휘 발달은 언어능력 전반의 기초를 형성하는 핵심 요소로, 이후의 인지, 사회적 상호작용, 의사소통 발달 등에 직·간접적인 영향을 미친다(Kim, 2014; Lee, 2019; Thal, 1991). 특히 18~36개월은 영유아가 폭발적으로 어휘를 습득하는 시기로(W. Park et al., 2023), 이 시기에 습득한 기초 어휘는 수용어휘와 표현어휘 모두에서 언어발달의 기반이 되며, 이후 학령기 언어능력 및 문해 발달과도 밀접하게 연결된다(Law et al., 2009; Rescorla, 2002). 따라서 이 시기의 아동이 어떤 어휘를 얼마나 습득하고 있는지를 파악하는 것은 언어재활, 아동교육, 부모상담 등 다양한 분야에서 매우 중요하다.

아동의 어휘 발달을 연구하는 방법은 전통적으로 세 가지 방식이 활용되어왔다. 먼저 간접 평가로써 부모를 통한 보고형 방식이 이루어지고 있다. 부모보고형 평가는 아동이 이해 및 표현할 수 있는 어휘를 부모나 주양육자가 판단하여 체크리스트 형태의 조사지에 응답하도록 하고 있다(W. Park et al., 2023). 부모보고형 평가는 사용이 용이하면서, 다양한 환경으로부터의 관찰 행동을 반영하여 대표성 있는 발달 자료를 얻을 수 있고, 평가 과정에 아동을 가장 잘 알고 있는 사람이 참여하였다는 점 등이 임상적 유용성을 가지고 있다(Thal et al., 2000). 한편, 많은 임상가들은 부모의 보고와 아동의 실제 수행력에서 차이를 언급하고 있는데, Dale(1991)은 부모보고의 타당성에 대하여 부모가 초기 언어학적 형태를 인식하는 능력이 부족하거나 자녀의 언어에 대하여 주관적 편향이 개입되어 아동을 실제보다 과대 또는 과소평가할 가능성이 있다고 하였다.

직접 평가 형태로는 임상가가 아동과 상호작용하며 반응을 평가하는 방식이 있다. 이 경우에 한 가지 방법은 표준화 어휘 검사도구를 활용하여 아동에게 그림 자극을 제공하면서 질문을 통해 아동이 어휘를 적절히 지적 또는 산출하는지 알아보는 것이다. 표준화 어휘 검사도구는 타당도와 신뢰도를 확보하고 있다는 장점이

있으나, 활용에 있어 아동의 어휘 목록들을 파악하는 것보다는 어휘 발달 수준을 또래와 비교하여 문제를 판별하는 데 주된 목적을 두고 있다(Kim et al., 2009). 이에 따라 특정 시점에서 대표적인 어휘를 기준으로 표준화가 이루어졌기 때문에 빠르게 변화하는 언어 환경과 사회적 맥락을 충분히 반영하지 못할 수 있다. 또한, 아동이 너무 어린 경우나 동반장애로 중증도가 심할 경우에는 구조화된 검사 수행이 어렵기도 하다(Kim, 2014).

또 다른 방식으로는 아동의 자발화를 분석하는 것이다. 이 평가는 임상가가 직접 평가의 형태로 상호작용하면서 아동의 자발화를 수집하거나, 아동의 일상을 일정 시간 동안 녹음 및 녹화하여 수집한 언어 샘플을 기록하여 분석하는 데 활용할 수 있다(Cha et al., 2014; Kim & Yun, 2017; Park & Lee, 2011). 자발화는 목적에 따라 의미·구문·화용 등 여러 가지 분석이 가능한데, 어휘 발달 연구에서는 주로 아동이 산출한 다른 낱말 수(number of different words: NDW), 전체 낱말 수(number of total words: NTW), 어휘 다양도(type-token ratio: TTR) 등을 분석하여 의미론적 발달에 참고하기도 한다(Kim, 2014). 아동의 자발화 수집은 실제적이고 생생한 언어 자료를 얻을 수 있다는 장점이 있지만, 연구자가 확보할 수 있는 표본의 규모와 대표성에 제한이 따른다. 그리고 임상가가 언어 샘플을 전사하고 분석하는 과정에 많은 시간이 소요되고 힘들다는 실제적인 어려움이 존재한다.

국내에서는 주로 직접 평가 방식의 표준화 어휘 검사도구로 수용·표현 어휘력 검사(Receptive and Expressive Vocabulary Test: REVT, Kim et al., 2009)를 사용하고 있다. 이 도구는 2세 6개월~16세 이상 성인을 대상으로 하기 때문에 폭넓은 연령의 어휘 평가에 활용도가 높지만, 36개월 이하 수준 영유아의 어휘 발달을 평가하기에는 제한점이 있다. 이에 36개월 이하 연령 수준 영유아의 어휘를 평가할 때는 부모보고형 방식의 표준화 어휘 검사도구인 한국판 맥아더-베이츠 의사소통발달 평가(Korean MacArthur-Bates Communicative Development Inventories: K M-B CDI, Pae & Kwak, 2011)가 대표적으로 활용되고 있다. 그러나 Kim 등(2020)은 시대의 변화에 따라 아동의 어휘 검사 목록이 수정되어야 할 필요성을 주장하며 추가·변경·제외 등의 재구성을 제안하였으며, W. Park 등(2023)도 K M-B CDI(Pae & Kwak, 2011)의 개발 시점이 10년이 지난 점 및 언어의 역동성을 언급하며 어휘 목록 개정의 필요성을 제기한 바 있다. 예를 들어, 부모보고를 통해 어휘를 평가할 시, 조사 목록에 컴퓨터, 휴대폰 등의 어휘가 없다면 이들 어휘는 중재 계획 시 고려되지 않을 확률이 발생한다. 이처럼 기존의 전통적인 평가 방법들은 연구 및 임상 현장에서 아동의 어휘 발달을 이해하는 데 기여했지만, 언어 환경의 변화를 실시간으로 반영하거나 대규모 데이터를 활용하는 데에는 한계점이 있다.

이와 관련하여 최근 인공지능(artificial intelligence: AI) 기술, 특히 생성형 인공지능(generative AI: GenAI)의 발전은 언어재활과 아동 언어 연구에도 새로운 가능성을 제시하고 있다. GenAI는 대규모 언어 데이터를 기반으로 어휘 사용의 패턴과 빈도를 분석할 수 있어, 특정 연령 아동에게서 나타나는 고빈도 어휘 목록을 빠르고 광범위하게 생성해낼 수 있다. 이는 기존의 시간과 비용이 많이 드는 언어 샘플 수집 과정을 보완하거나 표준화 어휘 검사도구를 확장하고 갱신하는 데에도 활용 가능성이 있다.

이러한 GenAI의 기술이 미국의 인공지능연구소 OpenAI에서 2022년 챗지피티(chat generative pre-trained transformer: ChatGPT)를 출시하면서 전문가들의 점유물이 아닌 누구나 일상에서 쉽게 접근하고 이용할 수 있는 기술이 되었다. 이에 따라 현재 언어재활 분야에서도 방대한 정보처리를 통한 생성 기술에 기반한 GenAI를 임상과 연구 현장에 적용하고자 하는 시도를 활발히 하고 있다(Kang et al., 2025; Lee & Yoon, 2025). Lee와 Yoon(2025)은 ChatGPT를 언어재활에 활용할 경우, 어휘력, 이야기 기술, 문해력 등의 언어재활 자료를 제작하는 데 보조적으로 사용할 수 있다고 하였다. 실제로 언어재활 분야에서 AI 기반 도구를 활용한 평가 및 중재 연구가 점차 증가하고 있으며(Cho et al., 2020; Hwang et al., 2020; Jung, 2020; Yun et al., 2020), 어휘 연구 또한 이러한 흐름 속에서 새로운 접근을 시도하고 활발히 적용할 필요가 있다.

그러나 GenAI가 생성한 어휘 자료를 그대로 임상이나 교육 현장에 적용하는 데에는 위험 요소가 존재한다. 특히 GenAI는 실제 근거가 없음에도 불구하고 사실과 다르거나, 그럴듯하지만 거짓된 정보를 생성하는 AI 할루시네이션(AI hallucination) 현상을 보인다(Park & Lee, 2024). 또한, National Evidence-based Healthcare Collaborating Agency(S. H. Park et al., 2023)에서는 AI는 블랙박스 특성, 제한된 일반화 등 고유한 특성이 있으므로, 이러한 요소들을 고려해서 AI 소프트웨어 기술의 의료현장 적용을 위해서 AI가 제공한 결과에 대해 평가 체계가 중요함을 강조한 바 있다. 예컨대, GenAI를 통해 어휘 목록을 생성할 경우, 디지털 미디어나 특정 온라인 환경에서 빈번히 등장하는 단어를 고빈도 어휘로 추천할 가능성도 있는데, 이러한 어휘가 실제 영유아의 언어발달에 적합하거나 발달적으로 중요한지는 보장되지 않는다. 또한, GenAI가 제시하는 자료는 데이터 출처의 특성에 따라 사회문화적 편향을 내포할 수 있어, 아동의 실제 언어 사용을 정확히 반영하지 못할 가능성도 있다. 따라서 GenAI 기반 어휘 자료를 신뢰성 있게 활용하기 위해서는 반드시 전문가 평가를 통한 타당도 검증 과정이 필요할 것이다.

이에 본 연구는 GenAI가 도출한 18~36개월 영유아 고빈도 어휘 목록의 타당성을 다각도로 검증하고자 한다. 구체적으로는 전문가 패널을 대상으로 발달 타당도, 기능 타당도, 빈도 타당도를 평가하여 GenAI의 생성어휘 목록의 적합성을 검증하고, 그 결과를 바탕으로 추출된 선정어휘 목록을 기존 표준화 어휘 검사도구와 비교하여 분석하고자 한다. 이러한 과정을 통해 아동 언어발달 연구와 언어재활 현장에 있어 GenAI 기반 어휘 자료의 활용 가능성과 한계를 탐색하고, 향후 어휘 평가 및 중재 도구 개발에 기초 자료를 제공하고자 한다.

II. 연구 방법

1. 연구 참여자

본 연구는 건양대학교 생명윤리위원회(Institutional Review Board: IRB)로부터 사전승인을 받은 후 실시되었다(KYU 2025-02-023-001). 전문가 조사의 경우, 전문가 패널 구성원의 선정이 중요한데, 전문 분야의 적절성, 전문적 지식과 능력, 구성원의 수 등을 고려해야 한다(Anderson, 1995; Dalkey et al., 1969; Kim, 2022). 이에 따라 전문가의 학위 전공 분야, 학력 및 경력 요건에 따라 다음과 같은 선정기준에 의해 연구 참여자를 모집하였다.

첫째, 전문가의 학위 전공 분야는 언어 및 아동 발달 관련 분야(언어치료학, 아동발달학, 유아교육학)를 전공한 자로, 둘째, 해당 분야의 석사 또는 박사학위 소지자면서, 셋째, 언어재활기관, 보육기관, 유아교육기관 등에 10년 이상 근무한 자로 정하였다. 기준을 충족한 전문가 패널 11명을 선정하였으며, 이력 정보는 Table 1에 제시하였다.

Table 1. Participants' information

Category		*n* (%)
Major	Speech-language pathology	9 (81.82)
	Child development	1 (9.09)
	Early childhood education	1 (9.09)
Education	Master's degree	4 (36.36)
	Doctor's degree	7 (63.64)
Career	10~15 years	2 (18.18)
	16~20 years	5 (45.45)
	21~25 years	1 (9.09)
	25~30 years	1 (9.09)
	31 years ≤	2 (18.18)
Total		11 (100.00)

2. 연구 도구

최근 국내에서는 OpenAI의 ChatGPT, Google의 Gemini, Microsoft의 MS Copliot, Naver의 HyperCLOVA X 등이 생성형 AI 서비스로 활용되고 있다. 이중 언어재활 전공자들이 가장 많이 사용한다고 보고된(Lee & Yoon, 2025), OpenAI의 ChatGPT를 연구 도구로 선정하였다.

ChatGPT-4.0을 활용하여 18~36개월 영유아의 고빈도 어휘 목록 500개를 생성하였다. 프롬프트는 "18~36개월 아이들이 많이 사용하는 어휘를 500개 추천해줘"라고 입력하였으며, 1회 시도에서 350개 어휘가 생성되어, "500개까지 추천해줘"를 재입력하여 추가 생성하였다. 생성된 어휘에 대하여 일차적으로 중복 및 어휘 범주의 적절성을 검토하였다. 그 결과, 500개 어휘 중 2.8% 비율인 14개 어휘가 중복으로 나타나 삭제하였고, 생성된 어휘의 범주는 100% 적절한 것으로 확인되었다.

이에 따라 검증할 생성어휘 목록은 사람 관련 범주 49개, 신체 범주 42개, 동물 범주 49개, 음식 범주 50개, 사물 범주 62개, 행동 범주 51개, 상태 범주 66개, 장소 범주 52개, 자연 범주 28개, 색깔 범주 22개, 숫자 범주 10개, 모양 범주 5개 등의 총 12개 범주와 486개 어휘 목록으로 구성되었다.

3. 연구 절차

생성어휘 목록 486개에 대하여 전문가 패널 11명이 참여하여 발달 타당도, 기능 타당도, 빈도 타당도를 판단하였다. 1~5점 리커트(Likert) 척도 기법에 따라 응답하도록 하였으며, 자유반응형 질문을 추가하여 검토 의견을 자유롭게 기술하도록 하였다. 전반적인 분석 과정은 다음과 같다.

첫째, 발달 타당도는 18~36개월 영유아의 수준에서 발달적으로 적절한지를 판단하도록 하였고, 둘째, 기능 타당도는 18~36개월 영유아의 환경에서 기능적으로 중요한지를 판단하도록 하였다. 셋째, 빈도 타당도는 18~36개월 영유아의 생활에서 사용 빈도가 높은지를 판단하도록 하였다.

다음으로 발달, 기능, 빈도 측면의 타당도 기준에 모두 적합한 어휘를 추출하여 선정하고, 전체 생성어휘 목록에서 어휘가 선정된 추출 비율을 분석하였다. 마지막으로 선정어휘를 기존의 표준화 어휘 검사도구인 K M-B CDI(Pae & Kwak, 2011)와 비교·분석하여 일치율을 산출하였다. 이상의 분석 절차를 Figure 1에 도식화하여 제시하였다.

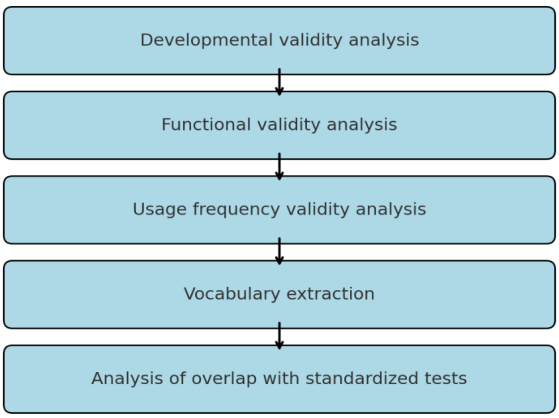

Figure 1. Analysis procedure

4. 자료 분석

응답 결과에 대한 자료 분석은 IBM SPSS Statistics 26과 Excel을 활용하였다. 내용 타당도 검증을 위한 분석 요소는 평균 점수, 내용 타당도 비율(content validity ratio: CVR), 합의도(agreement), 수렴도(convergence)를 산출하였다.

CVR은 내용이 타당하다고 응답한 전문가의 수가 50% 이상일 경우 (+)값, 타당하지 않다고 응답한 전문가의 수가 50% 이상일

경우 (-)값을 가지는데, 수치가 1에 가까울수록 내용 타당도가 높은 것으로 해석된다. 타당하다고 판단할 수 있는 CVR의 최소값에 대한 기준은 .59 이상을 적합 기준으로 하였다(Lawshe, 1975).

합의도는 전문가 간 어느 정도 합의가 이루어졌는가를 검증하는 것으로, 완전 합의했을 때 1의 값을 가지며, 의견 편차가 클 경우 수치가 감소한다. 반면, 수렴도는 전문가 조사를 통해 얻은 응답 결과가 한 점에서 모두 수렴했을 때 0의 값을 가지며, 의견 편차가 클 경우 수치가 증가한다. 즉, 합의도는 1에 가까울수록, 수렴도는 0에 가까울수록 전문가들의 의견이 합의가 이루어진 것으로 본다(Kang, 2008). 이에 따라 본 연구에서는 분석 요소의 적합 기준을 최종적으로 평균 3.00 이상, CVR .59 이상, 합의도 .70 이상, 수렴도 .50 이하를 기준으로 하였다(Table 2).

Table 2. Validity criteria

Types	Mean	CVR	Agreement	Convergence
Criteria	3.0 ≤	.59 ≤	.70 ≤	≤ .50

Note. CVR=content validity ratio.

선정어휘의 추출 단계에서는 이 기준에 모두 충족되는 어휘만 선정하였다. 즉, 미달되는 지표가 1개 이상 있는 어휘는 배제하였으며, 적합 기준을 충족하더라도 전문가들의 주관적 의견이 다수 있는 경우에 이를 종합적으로 반영하여 수정·보완하였다.

Ⅲ. 연구 결과

1. AI 기반 생성어휘의 내용 타당도

생성어휘 486개 목록의 내용 타당도를 발달, 기능, 빈도 측면에서 분석한 결과는 Table 3과 같다.

Table 3. Content validity analysis

Types	Mean	CVR	Agreement	Convergence
DV	4.24	.62	.80	.10
FV	4.22	.60	.80	.13
UFV	4.15	.59	.78	.11

Note. DV=developmental validity; FV=fuctional validity; UFV=usage frequency validity; CVR=content validity ratio.

구체적으로 살펴보면, 발달 타당도의 평균 점수는 5점 만점에 4.24점이었고, CVR=.62, 합의도=.80, 수렴도=.10으로 나타났다. 기능 타당도는 평균 4.22점이었으며, CVR=.60, 합의도=.80, 수렴도=.13으로 나타났다. 또한, 빈도 타당도는 평균 4.15점, CVR=.59, 합의도=.78, 수렴도=.11로 나타났다. 즉, 전체 어휘 목록의 발달, 기능, 빈도 타당도는 전반적으로 타당한 비율이 높았으며, 합의도 및 수렴도는 전문가들의 의견 일치가 이루어진 수준으로 나타났다.

2. 적합 어휘 추출 및 선정

타당도 기준을 충족하는 적합 어휘 목록을 추출한 결과, 486개 생성어휘 중 타당도 기준에 적합한 어휘는 313개로 선정되었다.

Table 4를 통해 12개 범주별로 살펴보면, 사람 어휘가 24개, 신체 어휘 28개, 동물 어휘 39개, 음식 어휘 42개, 사물 어휘 53개, 행동 어휘 48개, 상태 어휘 37개, 장소 어휘 15개, 자연 어휘 10개, 색깔 어휘 7개, 숫자 어휘 5개, 모양 어휘 5개 등이 선정되었다.

Table 4. Selection of valid vocabulary

Category	Selected vocabulary	Generated vocabulary	Selection ratio
	n (%)	*n* (%)	%
People	24 (7.67)	49 (10.08)	48.98
Body	28 (8.94)	42 (8.64)	66.67
Animals	39 (12.46)	49 (10.08)	79.59
Food	42 (13.42)	50 (10.29)	84.00
Objects	53 (16.93)	62 (12.76)	85.48
Actions	48 (15.34)	51 (10.49)	94.12
States	37 (11.82)	66 (13.58)	56.06
Places	15 (4.79)	52 (10.70)	28.85
Nature	10 (3.19)	28 (5.76)	35.71
Colors	7 (2.24)	22 (4.53)	31.82
Numbers	5 (1.60)	10 (2.05)	50.00
Shapes	5 (1.60)	5 (1.03)	100.00
Total	313 (100.00)	486 (100.00)	64.40

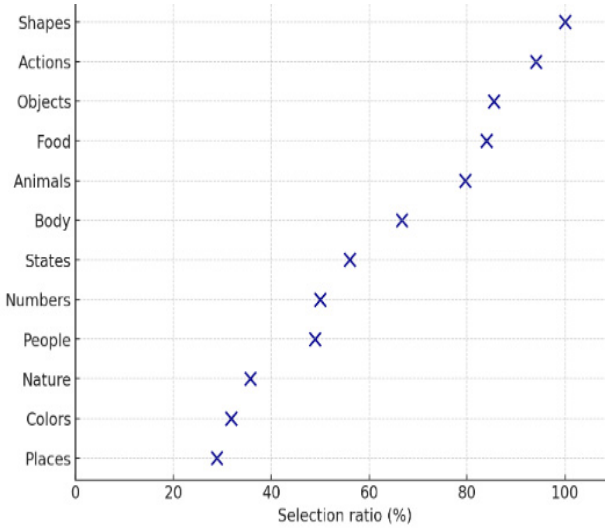

Figure 2. Vocabulary selection ratio

미선정 어휘의 예로는 사람 어휘에서 '기사, 사촌, 형수, 이웃, 선배, 후배, 조카, 형제, 자매, 친척', 신체 어휘에서 '피부, 심장,

속눈썹, 뺨, 종아리, 허벅지, 겨드랑이', 동물 어휘에서 '고슴도치, 수달, 까치, 까마귀, 무당벌레, 메뚜기, 지렁이', 음식 어휘에서 '두부, 버섯, 양파, 회', 사물 어휘에서 '킥보드, 공책, 밥솥, 전등', 행동 어휘에서 '건네다, 끌다', 상태 어휘에서 '지루하다, 답답하다, 우울하다, 설레다, 뿌듯하다, 부럽다, 의심하다, 긴장하다', 장소 어휘에서 '도서관, 시장, 백화점, 버스정류장, 공항, 캠핑장, 수족관, 미술관, 교실', 자연 어휘에서 '호수, 번개, 천둥, 안개, 이슬, 더위, 추위, 장마, 폭풍', 색깔 어휘에서 '주황, 보라, 갈색, 회색, 연두색, 남색', 숫자 어휘에서 '육, 칠, 팔, 구, 십' 등이 있었다.

전체 추출 비율은 64.40%였으며, 모양(100.00%), 행동(94.12%), 사물(85.48%), 음식(84.00%), 동물(79.59%), 신체(66.67%) 범주 순으로 적합한 비율이 높았다. 장소 범주는 추출 비율이 28.85%로 가장 낮게 나타났다. 범주별 추출 비율은 Figure 2에 점도표로 제시하였으며, 선정된 어휘의 목록 리스트는 Appendix 1에 제시하였다.

3. 선정어휘와 K M-B CDI 목록 비교

선정어휘 중에서 K M-B CDI(Pae & Kwak, 2011)의 어휘 목록과 일치하는 어휘의 비율을 산출하였다. 그 결과, 총 313개 선정어휘 중 241개 어휘가 일치하여 일치율 77.00%로 나타났다. K M-B CDI(Pae & Kwak, 2011)의 어휘 목록에는 없으나 GenAI 기반의 차별화된 어휘 72개는 Table 5와 같다.

Table 5. GenAI-specific vocabulary list

Category	Vocabulary
People	Older sister, kid
Body	Toe, fingernail, forehead, chest, elbow, palm, sole
Animals	Wolf, mole, dolphin, shark, whale, kangaroo, eagle, spider, cicada
Food	Rice, oriental melon, pear, cucumber, beef, pork, ham, sausage, chicken, seaweed
Objects	Colored pencil, police car, ambulance, traffic light, seat belt, computer, cell phone, remote control, air conditioner, desk, curtain, switch
Actions	Touch, chew, swallow, stack, organize, shake
States	Smell, warm, cool, angry, bright, excited, surprised, embarrassed, happy, annoyed, few, low
Places	Daycare center, cafe, pharmacy
Nature	Ice
Colors	Green, pink
Numbers	One, two, three, four, five
Shapes	Triangle, square, heart

범주별로 살펴보면, 사람 범주에서 '누나, 아이', 신체 범주에서 '발가락, 손톱, 이마, 가슴, 팔꿈치, 손바닥, 발바닥', 동물 범주에서 '늑대, 두더지, 돌고래, 상어, 고래, 캥거루, 독수리, 거미, 매미', 음식 범주에서 '밥, 참외, 배, 오이, 소고기, 돼지고기, 햄, 소시지, 치킨, 미역', 사물 범주에서 '색연필, 경찰차, 구급차, 신호등, 안전벨트, 컴퓨터, 휴대폰, 리모콘, 에어컨, 책상, 커튼, 스위치', '행동' 범주에서 '만지다, 씹다, 삼키다, 쌓다, 정리하다, 흔들다', 상태 범주에서 '냄새나다, 따뜻하다, 시원하다, 화나다, 밝다, 신나다, 놀라다, 부끄럽다, 행복하다, 짜증나다, 적다, 낮다', 장소 범주에서 '어린이집, 카페, 약국', 자연 범주에서 '얼음', 색깔 범주에서 '초록, 분홍', 숫자 범주에서 '일, 이, 삼, 사, 오', 모양 범주에서 '세모, 네모, 하트' 등이었다.

Ⅳ. 논의 및 결론

본 연구는 생성형 인공지능을 활용하여 도출된 18~36개월 영유아 고빈도 어휘 목록의 타당성을 검증하고, 이를 기존 표준화 어휘 검사도구와 비교·분석함으로써 언어재활 및 교육적 활용 가능성을 탐색하였다. 이러한 시도는 최근 AI 활용이 교육 및 의료 분야 전반에서 확산되는 흐름 속에서(Floridi & Chiriatti, 2020), 아동 어휘 발달 연구에도 새로운 가능성을 열었다는 점에서 의의가 있다. 본 연구의 결과를 요약하면 다음과 같다.

첫째, GenAI 기반 생성어휘 목록에 대한 전문가 타당도 평가에서 발달 타당도, 기능 타당도, 빈도 타당도 모두 평균 점수 4.00점 이상, CVR .59 이상, 합의도 .70 이상, 수렴도 .50 이하 수준으로 나타나 평균적으로 높은 타당도를 확보하였다(Kang, 2008; Lawshe, 1975). 이는 GenAI가 도출한 어휘가 전반적으로 적절하다는 전문가 합의를 보여준다.

둘째, 타당도의 적합 기준을 충족한 선정어휘는 총 313개였다. 이는 전체 생성어휘 486개 중 약 64.40%가 전문가 검증을 통과한 결과로, 대체로 타당한 어휘의 비율이 높다고 할 수 있다. 그러나 한편으로는 GenAI가 생성한 어휘를 그대로 언어재활 및 교육 현장에 적용하기에는 무리임을 시사한다. 따라서 실제 평가 및 중재에 활용하기 위해서는 전문가 합의와 체계적 선별을 통해 어휘 평가 목록의 타당성을 더 높여야 할 필요가 있다. 본 연구에서 전문가 집단 검증 과정에서 모양, 행동, 사물, 음식, 동물, 신체 범주는 적합 어휘로 추출된 비율이 높았지만, 일부 범주(예: 장소, 색깔, 자연, 사람 등)에서는 상대적으로 적은 비율의 어휘가 추출되었다. 주된 원인으로는 어휘의 범주는 적합하였지만, 난이도 수준이 영유아에게 높다고 판단되는 것이 많았다. 예를 들어, 사람 범주에서 선배, 후배, 이웃 등의 어휘가 생성되었는데, 이는 18~36개월 영유아에게 습득 근거가 매우 부족하므로 AI 할루시네이션(AI hallucination) 현상으로 추측된다. 또한, 장소 범주에서 캠핑장, 수족관, 미술관 등의 어휘가 생성되었는데 일상적 맥락에서 사용 빈도가 낮거나, AI 학습 데이터의 미디어 및 도시 중심 언어 편향에서 비롯되었을 가능성도 고려할 수 있다. 따라서 GenAI 기반 목록을 활용할 경우, 전문가의 검토와 보완을 병행하는 것이 필수적이다.

셋째, 최종 선정어휘 목록과 K M-B CDI(Pae & Kwak, 2011)의 목록 간 일치율은 약 77.00%로 나타났다. 이와 같은 결과는 GenAI 기반 어휘가 기존의 표준화 어휘 검사도구를 충분히 반영함과 동시에, GenAI가 최신 언어 환경을 반영하여 기존 도구

에 포함되지 않은 새로운 어휘를 제시할 수 있음을 보여준다. 예를 들어, '컴퓨터, 휴대폰, 에어컨, 카페' 등은 새로운 어휘의 추가 사례가 되며, '주방, 마트'는 기존 어휘 '부엌, 가게'의 대체 표현을 반영하였는데, 이러한 결과는 Kim 등(2020)의 제안과 일맥상통한다. 표준화 어휘 검사는 이러한 변화에 따라 수시로 최신화하기에 제한점이 있으므로, GenAI 기반의 어휘 생성 활용이 보완적 수단이 될 수 있을 것으로 생각된다.

본 연구의 결과는 다음과 같은 시사점을 갖는다. 우선 본 연구는 GenAI의 학문적 기여 가능성을 확인하였다. 기존 어휘 연구가 부모보고(Kim & Yun, 2017)나 자발화 샘플 수집(Park & Lee, 2011)에 의존했던 한계를 넘어, 대규모 언어 데이터를 기반으로 어휘를 추출·분석하고 전문가 검증을 거친다는 새로운 어휘 연구의 패러다임을 제시하였다.

또한, 본 연구에서 검증된 313개 선정어휘는 임상적·교육적 활용 가치가 높다. 전문가 합의를 거친 어휘는 언어발달 평가 도구, 중재 프로그램, 부모 상담 자료 등에 기초 자료로 적용할 수 있으며, 특히 변화하는 언어 환경 속 아동의 최근 어휘 사용을 반영한다는 점에서 교육적 의의가 크다.

이와 함께 본 연구는 기존 표준화 어휘 검사도구의 보완 가능성을 보여주었다. 현행 표준화 검사도구는 10년 이상 과거의 특정 시점 및 언어 환경에 맞추어 개발된 것이 많아(Kim et al., 2009; Pae & Kwak, 2011), 최근 디지털 환경과 사회적 변화가 충분히 반영되지 못하고 있다. 반면 GenAI 기반 어휘는 온라인 언어 환경을 포함한 최신 언어 패턴을 반영할 수 있어 전통적 도구와 상호 보완적으로 발전할 가능성이 있다.

그럼에도 불구하고 본 연구에는 몇 가지 제한점이 있다. 이를 논의하며 후속 연구를 제언하면 다음과 같다.

첫째, AI 학습 데이터의 출처 불투명성이 주요 제한점이다. 대규모 언어 모델은 방대한 데이터를 활용하지만 그 출처와 특성이 공개되지 않아, 데이터 편향 문제가 제기되어 왔다. 이에 AI 학습 데이터의 투명성 제고가 필요하다. 향후 연구에서는 아동 언어 코퍼스나 공개된 데이터셋을 기반으로 AI를 훈련하여 신뢰도를 높여야 한다.

둘째, 공인 타당도 검증 미비이다. 본 연구는 선정어휘와 K M-B CDI(Pae & Kwak, 2011)의 목록 간 일치율을 산출하였으나, 공인 타당도는 동일 집단을 대상으로 두 검사를 시행하고 점수 간 상관을 분석해야 한다. 이에 대규모 집단을 통한 공인 타당도 연구가 필요하다. 18~36개월 영유아의 부모 집단을 대상으로 GenAI 기반 어휘 검사와 기존 표준화 검사와의 점수 상관을 비교하는 연구가 요구된다.

셋째, 아동을 대상으로 한 직접 평가 연구가 제안된다. 본 연구의 선정어휘 목록을 직접 평가 형태의 그림 자극으로 제작하여 실제 아동의 수행력을 측정하거나 아동의 자발화 언어 샘플에서 선정어휘 목록이 산출되는 빈도를 분석한다면 실제적인 영유아 고빈도 어휘의 유효성을 검증할 수 있을 것이다. 이때, 아동의 어휘 발달은 사회문화적 배경에 따라 달라질 수 있으므로, 다양한 맥락에서 전국적인 대규모 표집 연구가 이루어질 필요가 있다.

본 연구는 GenAI를 활용하여 18~36개월 영유아의 고빈도 어휘 목록을 도출하고, 전문가 타당도 검증을 통해 GenAI 기반 어휘 자료의 신뢰성과 활용 가능성을 확인함으로써 기존 어휘 연구의 한계를 보완하고자 하였다. 그 결과, GenAI 기반 어휘 목록이 발달적·기능적·사용 빈도 측면에서 전문가 검증을 통해 타당성이 전반적으로 확보되었음을 보여주었으며, 최종적으로 313개 적합 어휘를 선정하였다.

이러한 결과는 GenAI가 디지털 환경과 사회·문화적 변화가 반영된 최신 어휘 목록을 빠르고 대규모로 도출할 수 있는 장점을 가지며, 기존의 부모보고, 언어 샘플 수집 등의 전통적 평가 방식이 가지는 시간과 자원 소모의 한계를 줄일 수 있는 대안적 접근법으로서의 가능성을 보여준다. 결론적으로 본 연구는 GenAI가 제시한 고빈도 어휘 목록이 일정 수준 이상의 타당성을 확보하고 있으며, 임상 언어 평가, 언어 중재 자료 개발, 유아 교육 콘텐츠 설계 등 다양한 영역에서 유용하게 활용될 수 있는 가능성을 확인하였다.

그러나 동시에 현시점 GenAI의 어휘 생성이 반드시 발달적 적절성과 의사소통 기능적 중요성을 보장하지는 않기 때문에, 내용 타당도 검증을 통한 신중한 필터링 과정이 필요하다는 점 또한 확인하였다. 나아가 GenAI 기반 자료와 전통적 검사도구가 상호 보완적으로 활용될 때, 아동 언어 평가와 중재의 정밀성이 크게 강화될 것으로 기대된다. 향후 연구에서는 다양한 언어 환경, 연령별 발달 수준, 지역·문화적 특성을 고려한 GenAI 기반 어휘 목록의 정교한 구성과 검증이 지속되어야 할 것이다. 이를 통해 언어발달 수준에 적합한 어휘 자료를 바탕으로 임상 현장에서는 보다 정밀한 언어 평가와 중재 계획 수립이 가능해지고, 교육 현장에서는 발달단계에 맞는 교수 자료 설계에 기여할 수 있을 것이다. 본 연구가 인공지능 기술을 영유아 어휘 발달 연구 및 실제 적용에 접목한 선도적 사례로서 기여하기를 바란다.

Reference

Anderson, D. R. (1995). *Strands of system: The philosophy of Charles Peirce.* West Lafayette, IN: Purdue University Press.

Cha, J.-E., Kim, J.-M., Kim, S.-J., Yoon, M.-S., & Chang, M.-S. (2014). Substantives in the vocabulary of typically developing young children. *Communication Sciences & Disorders, 19*(4), 430-446. doi:10.12963/csd.14191

Cho, J. Y., Chu, H. S., & Han, J. H. (2020). The effect of artificial intelligence speakers on the improvement of the vocabulary and information memory among teenagers and adults with hard intellectual disabilities. *Proceedings of 21st Conference on the Korean Speech-Language and Hearing Association*, 302-308.

Cho, K. (2018). *Developing a career decision-making status inventory for university students in Korea* (Doctoral dissertation). Seoul National University, Seoul.

Dale, P. S. (1991). The validity of a parent report measure of vocabulary and syntax at 24 months. *Journal of Speech,*

Language, and Hearing Research, 34(3), 565-571. doi:10.1044/jshr.3403.565

Dalkey, N. C., Brown, B. B., & Cochran, S. (1969). *The Delphi method III: Use of self ratings to improve group estimates.* Santa Monica, CA: Rand.

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines, 30*(4), 681-694. doi:10.1007/s11023-020-09548-1

Hwang, D. J., Pyo, S. M., & Kim, B. A. (2020). The effects of repetitive reading intervention through artificial intelligence on reading fluency of children with language learning disabilities. *Proceedings of 21st Conference on the Korean Speech-Language & Hearing Association,* 292-297.

Jung, S.-I. (2020). Case study on intervention of conversational skills in children with autism spectrum disorder using AI speaker. *Proceedings of 21st Conference on the Korean Speech-Language & Hearing Association,* 309-314.

Kang, M. G., Lee, S. B., Cho, E. B., NA, D., & Yoon, J. H. (2025). Exploring the speech transcription and analysis capabilities of speech recognition and generative AI: Focusing on the utterances of older adults. *Journal of Speech-Language & Hearing Disorders, 34*(3), 1-11. doi:10.15724/jslhd.2025.34.3.001

Kang, Y. J. (2008). *Understanding and application cases of the Delphi technique.* Seongnam: Korea Employment Agency for the Disabled Employment Development Institute.

Kim, S. H. (2022). *A study on the characteristics of children with nonspecific language impairment through the development and application of inference ability evaluation tasks* (Doctoral dissertation). Daegu University, Gyeongbuk.

Kim, Y. M., & Yun, E. (2017). Exploring the meanings of usage about onomatopoeia and mimetic words in a child play situation. *Korean Journal of Child Studies, 38*(1), 63-76. doi:10.5723/kjcs.2017.38.1.63

Kim, W., Lee, J., Seo, D., & Yang, H. (2020) Suggestions for composing an 18-36 month infant vocabulary checklist. *Korea Journal of Child Care & Education, 123,* 1-25. doi:10.37918/kce.2020.07.123.1

Kim, Y. T. (2014). *Assessment and treatment of language disorders in children* (2nd ed.). Seoul: Hakjisa.

Kim, Y. T., Hong, K. H., Kim, K. H., Jang, H. S., & Lee, J. Y. (2009). *Receptive & Expressive Vocabulary Test* (REVT). Seoul: Seoul Community Rehabilitation Center.

Law, J., Rush, R., Schoon, I., & Parsons, S. (2009). Modeling developmental language difficulties from school entry into adulthood: Literacy, mental health, and employment outcomes. *Journal of Speech, Language, and Hearing Research, 52*(6), 1401-1416. doi:10.1044/1092-4388(2009/08-0142)

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*(4), 563-575. doi:10.1111/j.1744-6570.1975.tb01393.x

Lee, S. B., & Yoon, J. H. (2025). Current applications and perceptions of generative AI in speech-language pathology clinical practice and research. *Journal of Speech-Language & Hearing Disorders, 34*(1), 155-169. doi:10.15724/jslhd.2025.34.1.155

Lee, Y. K. (2019). *Communication disorders in infants and toddlers.* Seoul: Hakjisa.

Pae, S., & Kwak, K. (2011). *Korean MacArthur-Bates Communicative Development Inventories* (K M-B CDI). Seoul: Mindpress.

Park, D.-M., & Lee, H.-J. (2024). Literature review of AI hallucination research since the advent of ChatGPT: Focusing on papers from arXiv. *Informatization Policy, 31*(2), 3-38. doi:10.22693/NIAIP.2024.31.2.003

Park, M. H., & Lee, J. (2011). A study for characteristics of vocabulary in spontaneous language sample aged from 3 to 5. *The Study of Education for Hearing-Language Impairments, 2*(1), 1-13. doi:10.24009/ksehli.2011.2.1.001

Park, S. H., Sul, A.-R., Park, S. J., Jhang, H., Kim, Y. H., & Choi, S. R. (2023). *A study to develop methods to assess AI-based medical software devices for adoption in real-world practice.* Seoul: National Evidence-Based Healthcare Collaborating Agency.

Park, W., Seok, H.-E., & Yim, D. (2023). Investigating the dynamics in 18-36 months toddlers' vocabulary acquisition. *Communication & Science Disorders, 28*(2), 197-210. doi:10.12963/csd.23970

Rescorla, L. (2002). Language and reading outcomes to age 9 in late-talking toddlers. *Journal of Speech, Language, and Hearing Research, 45*(2), 360-371. doi:10.1044/1092-4388(2002/028)

Thal, D., Jackson-Maldonado, D., & Acosta, D. (2000). Validity of a parent-report measure of vocabulary and grammar for Spanish-speaking toddlers. *Journal of Speech, Language, and Hearing Research, 43*(5), 1087-1100. doi:10.1044/jslhr.4305.1087

Thal, D. J. (1991). Language and cognition in normal and late-talking toddlers. *Topics in Language Disorders, 11*(4), 33-42. doi:10.1097/00011363-199111040-00006

Yun, E. M., Choi, H. S., Lee, J. K., & Kang, J. S. (2020). Reading intervention in children with reading difficulties through vocabulary improvement using Kakaomini. *Proceedings of 21st Conference on the Korean Speech-Language & Hearing Association,* 333-335.

Appendix 1. Selected vocabulary list

Category	No.	Word	Category	No.	Word	Category	No.	Word	Category	No.	Word
사람 (24)	1	엄마	음식 (42)	81	악어	행동 (48)	161	우산	상태 (37)	241	아프다
	2	아빠		82	하마		162	시계		242	졸리다
	3	할머니		83	캥거루		163	거울		243	화나다
	4	할아버지		84	참새		164	텔레비전		244	무섭다
	5	형		85	독수리		165	컴퓨터		245	배고프다
	6	누나		86	부엉이		166	휴대폰		246	목마르다
	7	오빠		87	벌		167	리모컨		247	맛있다
	8	언니		88	나비		168	냉장고		248	밝다
	9	동생		89	개미		169	에어컨		249	어둡다
	10	친구		90	거미		170	책상		250	시끄럽다
	11	선생님		91	매미		171	의자		251	조용하다
	12	아저씨		92	밥		172	침대		252	신나다
	13	아줌마		93	국		173	소파		253	재미있다
	14	이모		94	김치		174	베개		254	예쁘다
	15	삼촌		95	떡		175	이불		255	멋있다
	16	고모		96	빵		176	창문		256	귀엽다
	17	아기/아가		97	우유		177	문		257	좋다
	18	나		98	치즈		178	커튼		258	나쁘다
	19	너		99	계란/달걀		179	스위치		259	놀라다
	20	우리		100	사과		180	빗		260	부끄럽다
	21	아이		101	바나나		181	칫솔		261	행복하다
	22	어른		102	딸기		182	치약		262	짜증나다
	23	의사		103	포도		183	휴지		263	힘들다
	24	경찰		104	귤		184	수건		264	심심하다
신체 (28)	25	얼굴		105	수박		185	로봇		265	크다
	26	눈		106	참외		186	장난감		266	작다
	27	코		107	배		187	먹다		267	많다
	28	입		108	복숭아		188	마시다		268	적다
	29	귀		109	오렌지		189	(잠)자다		269	높다
	30	이/이빨		110	감		190	일어나다		270	낮다
	31	혀		111	토마토		191	씻다		271	길다
	32	손		112	고구마		192	가다	장소 (15)	272	집
	33	발		113	당근		193	오다		273	아파트
	34	팔		114	오이		194	앉다		274	어린이집
	35	다리		115	옥수수		195	서다		275	방
	36	손가락		116	고기		196	눕다		276	(키즈)카페
	37	발가락		117	닭고기		197	뛰다		277	화장실
	38	무릎		118	소고기		198	달리다		278	주방/부엌
	39	배		119	돼지고기		199	걷다		279	학교
	40	등		120	햄		200	놀다		280	유치원
	41	목		121	소시지		201	보다		281	공원
	42	어깨		122	라면		202	듣다		282	마트/가게
	43	엉덩이		123	피자		203	만지다		283	병원
	44	손톱		124	치킨		204	말하다		284	약국
	45	머리		125	과자/까까		205	웃다		285	수영장
	46	배꼽		126	아이스크림		206	울다		286	회사
	47	이마		127	초콜릿		207	던지다	자연 (10)	287	하늘
	48	입술		128	사탕		208	잡다		288	바다
	49	가슴		129	주스		209	밀다		289	강
	50	팔꿈치		130	물		210	당기다		290	산
	51	손바닥		131	콜라		211	열다		291	해/햇빛
	52	발바닥		132	요구르트		212	닫다		292	달
동물 (39)	53	강아지/개		133	미역		213	타다		293	바람
	54	고양이	사물 (53)	134	책		214	오르다		294	비
	55	토끼		135	공		215	내리다		295	눈
	56	병아리		136	인형		216	씹다		296	얼음
	57	닭		137	차/자동차		217	삼키다	색깔 (7)	297	빨강
	58	오리		138	기차		218	뱉다		298	파랑
	59	소		139	비행기		219	닦다		299	노랑
	60	돼지		140	배		220	(글)쓰다		300	초록
	61	양		141	자전거		221	읽다		301	검정
	62	말		142	연필		222	그리다		302	흰/하얀
	63	코끼리		143	색연필		223	만들다		303	분홍
	64	사자		144	오토바이		224	쌓다	숫자 (5)	304	일
	65	호랑이		145	버스		225	부수다		305	이
	66	곰		146	트럭		226	정리하다		306	삼
	67	기린		147	헬리콥터		227	숨다		307	사
	68	원숭이		148	소방차		228	찾다		308	오
	69	여우		149	경찰차		229	입다	모양 (5)	309	동그라미
	70	늑대		150	구급차		230	벗다		310	세모
	71	다람쥐		151	신호등		231	흔들다		311	네모
	72	두더지		152	안전벨트		232	돌리다		312	하트
	73	펭귄		153	크레파스		233	넣다		313	별
	74	돌고래		154	가위		234	주다			
	75	상어		155	가방		235	냄새나다			
	76	고래		156	신발		236	따뜻하다			
	77	물고기		157	양말		237	시원하다			
	78	거북이		158	옷		238	뜨겁다			
	79	개구리		159	모자		239	차갑다			
	80	뱀		160	장갑		240	무겁다			

언어치료연구 2025년 제34권 제4호 043-051
ISSN 1226-587X / eISSN 2671-7158
JSLHD (Journal of Speech-Language & Hearing Disorders)

ORIGINAL ARTICLE
https://doi.org/10.15724/jslhd.2025.34.4.043
http://jslhd.org/

생성형 인공지능 기반 영유아 고빈도 어휘의 타당성 연구

김시현[1*]

[1] 건양대학교 보건복지대학원 언어치료학과 교수

목적: 본 연구의 목적은 생성형 인공지능을 활용하여 도출된 영유아 고빈도 어휘 목록의 타당성을 다각적으로 살펴보고, 어휘 발달 연구의 새로운 접근법으로 인공지능의 활용 가능성을 탐색하는 데 있다.

방법: ChatGPT-4.0을 통해 18~36개월 영유아의 고빈도 어휘 목록을 생성하고, 전문가 11명이 참여하여 발달, 기능, 빈도 측면에서 타당도를 평가하였다. 이를 타당도 기준에 따라 적합한 어휘를 선정하였으며, 선정된 추출 비율을 분석하였다. 또한, 선정어휘와 국내 표준화 어휘 검사도구인 한국판 맥아더-베이츠 의사소통발달 평가(K M-B CDI)의 목록을 비교하여 일치율을 분석하였다.

결과: 첫째, ChatGPT-4.0에 기반한 생성어휘 486개 목록은 전반적으로 발달, 기능, 빈도 측면에서 적합 기준을 충족하였다. 둘째, 기준을 충족하는 적합 어휘는 313개 어휘가 선정되었으며, 전체 목록의 64.40%에 해당되었다. 사람 어휘가 24개, 신체 어휘 28개, 동물 어휘 39개, 음식 어휘 42개, 사물 어휘 53개, 행동 어휘 48개, 상태 어휘 37개, 장소 어휘 15개, 자연 어휘 10개, 색깔 어휘 7개, 숫자 어휘 5개, 모양 어휘 5개 등이 선정되었다. 셋째, 선정어휘와 K M-B CDI 목록과의 일치율은 77.00%로 나타났다.

결론: 이러한 결과는 AI 기반 어휘 목록이 기존 표준화 검사도구의 범위를 보완하면서도, 변화하는 언어 환경을 반영할 수 있음을 시사한다. 본 연구는 AI를 활용한 어휘 발달 연구의 가능성을 제시하였으며, 영유아 교육 및 언어재활 현장에서 평가와 중재 자료로 참고될 수 있는 기초자료를 마련하였다.

검색어: 생성형 인공지능, 영유아, 어휘, 타당성

교신저자 : 김시현(건양대학교)
전자메일 : shkim1881@konyang.ac.kr

게재신청일 : 2025. 09. 11
수정제출일 : 2025. 09. 24
게재확정일 : 2025. 10. 31

이 논문은 2024년도 상반기 건양대학교 학술연구비 지원에 의하여 이루어진 것임.

ORCID
김시현
https://orcid.org/0000-0002-2476-4529

참고문헌

강민국, 이소빈, 조은별, 나덕렬, 윤지혜 (2025). 음성인식 및 생성형 AI의 발화 전사 및 분석 능력 탐색: 장노년층 발화를 중심으로. **언어치료연구, 34**(3), 1-11.

강용주 (2008). **델파이 기법의 이해와 적용 사례.** 성남: 한국장애인고용공단 고용개발원.

김영태 (2014). **아동언어장애의 진단 및 치료**(2판). 서울: 학지사.

김시현 (2022). **추론능력 평가과제 개발 및 적용을 통한 비특정 언어장애 아동의 특성 연구.** 대구대학교 대학원 박사학위 논문.

김영태, 홍경훈, 김경희, 장혜성, 이주연 (2009). **수용·표현 어휘력 검사.** 서울: 서울장애인종합복지관.

김윤미, 윤은주 (2017). 유아 놀이에서 의성어 · 의태어 사용 의미 탐구. **아동학회지, 38**(1), 63-76.

김화수, 이지우, 서다희, 양한나 (2020). 18-36개월 영아의 어휘검사 목록 구성을 위한 제안. **한국영유아보육학, 123,** 1-25.

박대민, 이한종 (2024). 챗 GPT 등장 이후 인공지능 환각 연구의 문헌 검토: 아카이브(arXiv)의 논문을 중심으로. **정보화정책, 31**(2), 3-38.

박미혜, 이전아 (2011). 자발화 표본에 나타난 3-5세 아동의 어휘 특성. **한국청각언어장애교육연구, 2**(1), 1-13.

박성호, 설아람, 박소진, 장호열, 김영희, 최서린 (2023). **AI 소프트웨어 의료기기의 의료현장 적용을 위한 평가방안 마련 연구.** 서울: 한국보건의료연구원.

박원정, 석혜은, 임동선 (2023). 18-36개월 아동의 어휘 습득 역동성 분석 연구. **Communication & Science Disorders, 28**(2), 197-210.

배소영, 곽금주 (2011). **한국판 맥아더-베이츠 의사소통 발달평가.** 서울: 마인드프레스.

윤은미, 최효선, 이진국, 강진석 (2020). 카카오미니를 활용한 어휘능력 향상을 통한 읽기중재. **제21회 한국언어치료학회 학술대회 발표논문집,** 333-335.

이소빈, 윤지혜 (2025). 언어병리 임상과 연구에서의 생성형 AI 활용 현황 및 인식. **언어치료연구, 34**(1), 155-169.

이윤경 (2019). **영유아 의사소통장애 발달, 평가, 중재.** 서울: 학지사.

정상임 (2020). 인공지능 스피커를 사용한 자폐스펙트럼장애 아동의 대화 기술 중재 사례 연구. **제21회 한국언어치료학회 학술대회 발표논문집,** 309-314.

조정예, 추호성, 한주희 (2020). 인공지능 스피커가 경도지적장애 청소년과 성인의 어휘인출 및 정보기억 향상에 미치는 효과. **제21회 한국언어치료학회 학술대회 발표논문집,** 302-308.

차재은, 김정미, 김수진, 윤미선, 장문수 (2014). 2-5세 일반 아동의 어휘 발달: 체언. **Communication Sciences & Disorders, 19**(4), 430-446.

황동준, 표승민, 김보애 (2020). 인공지능 기기를 통한 반복 읽기 중재가 언어학습장애 아동의 읽기 유창성에 미치는 영향. **제21회 한국언어치료학회 학술대회 발표논문집,** 292-297.