영어어문교육 *제29권* 3호 2023 가을
https://doi.org/10.35828/etak.2023.29.3.113

수준별 영어 수업에서의 정의적 영역에 대한 효과 분석*

윤지여**
(정동초등학교)
박옥희***
(배재대학교)

Yun, Jiyeo, & Park, OkHee. (2023). An analysis of effects on affective domain in level differentiated English classes. ***English Language & Literature Teaching 29*****(3), 113-133.**

As a follow-up study for meta-analysis on cognitive domain (Park & Yun, 2019), this study aims to investigate the overall effect on affective domain in level differentiated English classes. The procedure of meta-analysis consists of literature search with criteria for inclusion as well as data analysis including data coding and cleaning. After heterogeneity tests, random-effect model, subgroup analyses, and publication bias are conducted using Comprehensive Meta-Analysis software. Moderators used for subgroup analyses were; affective domains, schools, grades, and teaching methods and periods of level differentiated English classes. The results indicate that affective domain in level differentiated English classes had an overall mean effect size ($\bar{d}$ = 0.33) showing lower effect size than in cognitive domain ($\bar{d}$ = 0.74). More specific findings of meta-analysis are as follows: First, grouping factors based on affective domain had medium effects in confidence ($\bar{d}$ = 0.68), interest ($\bar{d}$ = 0.57) and participation ($\bar{d}$ = 0.51). Second, according to grade and to students in elementary school through university, the effects were larger with higher grade elementary school students and high school freshmen students. Third, among teaching methods, affective domain showed a large effect in extra supplementary classes. Lastly, the pedagogical implication and suggestions were further discussed.

[meta-analysis/affective domains/level differentiated English classes/effect size, 메타 분석/정의적 영역/수준별 영어 수업/효과 크기]

* 본 논문은 2023학년도 배재대학교 교내학술연구비 지원에 의하여 수행됨.
** 제1저자 *** 교신저자

Ⅰ. 서론

수준별 교육은 학습자들의 수준, 능력, 적성에 따른 개인차를 고려하여 교육 목표, 교육내용, 교육 방법 등을 달리하는 학습자 중심의 교육과정이다(Brown, 2007). 우리나라의 수준별 교육 현황을 보면, 2012년 기준 국내 대학의 70% 정도가 수준별 수업을 진행하고 있다고 보고되었다(권선희, 2012). 초중등학교에서는 제7차 교육과정 당시 수준별 수업을 전면 시행하였다가 제7차 개정 교육과정부터는 학교장의 재량에 따르는 것으로 바뀌었다(심지현, 임현우, 2017). 공교육에서의 이러한 정책 변경은 수준별 교육이 지향하는 교육 철학과 달리 수준별 수업의 효과에 대한 논쟁에서 비롯된 것이라 볼 수 있고, 이는 해외에서도 비슷한 상황을 보인다. 즉, 수준별 교육은 학습 효율성을 높이고 교육의 질을 높인다는 긍정적인 주장(Kulick, 1992; Slavin, 1987)과 수준별 교육은 하위 학습자들에게 불리하고, 학력 격차만 키운다는 부정적인 주장(Domina 외, 2017; Donaldson 외, 2017; Gamoran & Mare, 1989)이 팽팽하게 맞서고 있다(박옥희, 윤지여, 2019).

이러한 논쟁에 대해 박옥희와 윤지여(2019)는 국내에서 실시되고 있는 수준별 영어 수업에서 인지적 영역에 어떠한 효과가 있는지 메타 분석을 통해 알아보았다. 국내의 수준별 영어 수업에 대한 연구를 모아 메타 분석을 한 결과, 학업 성취도의 평균 효과 크기가 0.74로 매우 높게 나왔다. 수준별 영어 수업이 인지적 영역에서 효과가 있다는 결과와 마찬가지로 정의적 영역에서도 효과 있는지, 그리고 효과가 있다면 얼마나 효과가 있는지 궁금해졌다. 그래서 수준별 수업에서 인지적 영역의 효과 크기를 알아본 박옥희와 윤지여(2019)의 후속 연구로, 수준별 영어 수업의 정의적 영역에 대한 효과를 알아보는 것이 본 연구의 목적이다. 다시 말해, 우리나라 학습자를 대상으로 수준별 영어 수업을 일정 기간 시행한 뒤 학습자의 정의적 영역에서의 효과를 알아본 연구를 모아 효과 크기를 살펴보는 것이다. 구체적으로는 정의적 영역(흥미도, 자신감, 참여도, 만족도, 불안감, 동기, 태도)과 학년(초등학생, 중학생, 고등학생, 대학생), 수준별 수업 기간(10주에서 40주), 수준별 수업 형태(수준별 이동 수업, 학급 내에서 심화 보충 활동, 수준별 자료 활용 수업)에 따라 효과 크기가 어떠한지를 조사하고자 한다. 구체적인 연구 문제는 두 가지이다.

첫째, 수준별 영어 수업이 학습자에게 미치는 정의적 영역의 효과 크기는 어느 정도인가?

둘째, 수준별 영어 수업의 조절변인(moderator)인 ‘정의적 영역’, ‘학교’, ‘학년’, ‘수준별 수업 형태’, ‘수준별 수업 기간’에 따라 학습자에게 미치는 효과 크기는 어느 정도인가?

Ⅱ. 이론적 배경

1. 수준별 수업(differentiated classes)

수준별 수업의 이론적 근거는 학습자 중심 교육과정에서 시작되었다고 할 수 있다. '학습자 중심'이란 용어는 Rogers(1967)의 인본주의 심리학 또는 진보주의 교육관에서 비롯되었다. Rogers(1967)는 학습 과정에서 학습자의 역할을 강조하면서 교사는 학습자의 보조자 및 촉진자 역할을 해야 한다고 주장하였다. 뒤를 이어 Slavin(1987)은 성공적인 수준별 수업을 만들기 위해서 교수자는 학습자 개개인의 수준에 맞는 다양한 교육적 처방을 제공해야 한다고 하였다. 즉, 학습자의 개별적인 학습능력에 따라 교과과정을 편성하고, 차등적으로 학습 내용을 적용하며, 수준을 정할 때 학습자의 언어능력을 고려해야 한다고 하였다(박지영 외, 2012). Tomlinson(2014)은 수준별 수업에서 고려해야 할 내용을 다음과 같이 정리하였다. 첫째, '무엇을 차별화할 것인가(Differentiate what?)'로 교과내용, 교육과정 또는 학습 환경의 차별화가 필요하다. 둘째, '어떻게 차별화할 것인가(Differentiate how?)'로 학습자의 특성, 흥미, 개인 경험 등을 반영하는 것이 중요하다. 셋째, '왜 차별화할 것인가(Differentiate why?)'로 학습의 수월성, 학습 동기 부여 및 학습의 효율성을 고민해야 한다(박옥희, 윤지여, 2019).

국내에서의 수준별 영어 수업에 대한 연구 경향을 보면, 수준별 수업의 운영 방식과 실태 조사(김영현 2002; 남지영, 2020; 박상옥, 2004; 주미란, 2021), 학습자들의 인식 및 정의적 영역에 대한 조사(김영숙, 이지연, 2009; 김윤정, 2015; 박지영 외, 2012; 서정아, 2018; 안애랑, 박은수, 2014; 양정임, 2019; 이경화 외 2016; 임정완, 2014), 그리고 학업 성취도 조사(김은주, 성귀복, 2016; 박선호, 송현진, 2013; 이지은, 장종득, 2014; 하명호, 2014; 홍은화, 김경자, 2014) 등으로 나누어 볼 수 있다.

수준별 영어 수업에서의 정의적 영역에 관한 선행연구를 중심으로 살펴보면, 인지적 영역에 관한 연구와 마찬가지로 정의적 영역에 관한 연구 역시 수준별 수업을 위해 구성한 학습자 수준에 따라 어떤 차이와 의미가 있는지에 초점이 맞춰져 있고, 연구 결과 역시 다르게 보고되고 있다. 먼저, 수준별 수업이 하위반 학습자의 정의적 영역에 긍정적이라는 연구들이 있다. 김윤정(2015)은 서울 소재의 4년제 대학에서 수준별 교양 수업에 대한 효과와 만족도를 조사하였는데 기초반의 만족도가 제일 높았고, 중급반의 만족도가 제일 낮았다. 김영숙과 이지연(2009)은 대학 교양 영어 수강생들을 대상으로 만족도를 조사한 결과, 초급반 학생들의 만족도가 고급반 학생들의 만족도 보다 높게 나왔다. 박지영 외(2012)는 수준별 교양 영어 과목에 대한 만족도를 조사하였는데 초급반 학생들의 만족도가 제일 높았고, 고급반, 중위반의 순으로 그 뒤를 이었다. 안애랑과 박은수(2014)는 중학교 하위반 학생들을 대상으로 정의적 영역을 조사한 결과, 하위반 학생들의 학습 태도, 흥미도, 자신감이 높아졌고, 불안감은 낮아진 것으

로 조사되었다. 양정임(2019)은 대학 교양 영어 수업을 수강하는 신입생들을 대상으로 수준별 수업에 대한 인식 및 태도를 조사하였는데 초급반에서는 수준별 수업 선호도가 향상되었지만, 중급반과 상급반에서는 선호도가 낮아졌다는 결과를 얻었다.

반면, 수준별 수업은 하위반 학습자들의 정의적 영역에 부정적이라는 결과를 보고한 연구들도 있다. 이경화 외(2016)는 수준별 이동수업을 하는 초등학교 5학년과 6학년 학생들을 대상으로 정의적 영역에 관한 설문조사를 실시하였는데 하위반 학생들이 상위반 학생들보다 불안감이 더 높고, 태도와 흥미, 자신감에서는 더 낮은 결과를 보였다. 임정완(2014)은 고등학교 1학년 학생들을 대상으로 수준별 이동 수업에 대한 인식을 조사한 결과, 중위반과 하위반 학생들은 부정적인 인식이 많은 반면 상위반 학생들은 전반적으로 긍정적인 인식을 보여주었다. 한편, 서정아(2018)의 연구에서는 수준별로 교양 영어를 수강하는 학습자들이 상중하 모든 반에서 전반적으로 높은 만족도를 보였다.

이렇듯 선행연구는 수준별 수업이 정의적 측면에서 학습자 집단에 따라 효과가 서로 다르다는 결과를 보고하고 있다. 메타분석은 이렇게 상이한 결과를 가진 연구들을 모아 종합 분석하여 결론을 도출해 내는 하나의 연구방법이다. 따라서 메타분석을 통해 수준별 수업이 정의적 영역에 효과가 있는지, 효과가 있다면 어떤 부분인지 등에 관한 정확하고 구체적인 분석이 필요하다.

2. 정의적 영역(affective domain)

외국어 학습에서 정의적 요인(affective factors)은 교수학습의 촉진 혹은 저해 요소가 되어 외국어 학습의 성공이나 실패에 영향을 미칠 수 있다(Brown, 2010). 정의적 요인에 관한 주요 연구로 학습자의 태도와 동기를 강조한 Gardner와 Lambert(1972)의 학습 이론과 Krashen(1982)의 정의적 여과 가설(affective filter hypothesis)을 들 수 있다. 특히, Krashen(1982)은 외국어 학습에서 정의적 여과 장치는 언어 입력(input)이 이루어지는 중요한 요소라고 하였다. 이 여과 장치에 영향을 주는 정의적 요인으로 학습자의 태도, 신념, 자기 확신, 불안, 문화 충격 등을 들었다. 또한 외국어 습득의 성공과 정의적 영역의 상관관계에서 불안감이 낮을수록, 동기가 높을수록, 자존감과 자신감이 클수록 성공적인 외국어 습득이 가능하다고 하였다(박옥희, 2012). 학자에 따라 정의적 요인의 분류가 조금씩 다르고 개념이 서로 중복되기도 하는데, 박옥희(2012)는 표 1과 같이 정리하였다.

구체적으로 흥미에 관하여 Krashen(1983)은 외국어 학습 과정에서 학습자가 갖는 흥미는 외국어를 배울 수 있는 지속적인 학습활동이 될 수 있고, 학습의 촉진제 역할을 할 수 있다고 하였다. Gardner(1985)는 동기를 목표언어의 문화권에 통합되고 그 문화권의 사회적 일원이 되려고 하는 통합적 동기(integrative motivation)와 구직과 같은 현실적인 목표를 위해 외국어를 배우는 도구적 동기

(instrumental motivation)로 분류하였다. Deci(1975)는 외적인 보상과는 상관없는 자기만족을 위한 내적 동기(intrinsic motivation)와 외부로부터의 보상을 위한 외적 동기(extrinsic motivation)로 분류하였다.

표 1
정의적 영역의 분류와 특징(박옥희, 2012, p. 428)

정의적 영역	특징
흥미(interest)	좋아하는 대상이나 활동에 더 많은 관심과 노력을 투자하여 뛰어나거나 인정받으려는 심리적 특성
동기(motivation)	학습자의 강한 의지가 학습과정에 반영되는 것
불안(anxiety)	근심, 좌절, 우려, 걱정, 두려움 같은 감정상태
자신감(self- confidence)	자아존중(self-esteem)과 자기 능력에 대한 신뢰
태도(attitude)	사람이나 사물에 대한 개인의 심리적 반응 경향
참여(participation)	학습자들이 지적 추측이나 언어적 감각을 통해 실수의 가능성에 대한 위험을 감수하면서까지 학습에 적극적으로 대응하는 것
만족(satisfaction)	학습자가 학업 성취에 대해 긍정적인 느낌을 가지는 것

Oxford(1999)는 불안을 학습에 도움이 되는 촉진적 불안(facilitative anxiety)과 방해가 되는 저하적 불안(debilitative anxiety)으로 분류하였다. Horwitz 외(1986)는 외국어 학습 상황에서 발생할 수 있는 불안을 외국어 불안(foreign language anxiety)이라고 하였는데 이를 세 가지로 설명하였다. 첫째, 원어민과 외국어로 의사소통할 수 있는 자기 능력에 대한 의심에서 오는 불안감, 둘째, 자신이 사용한 외국어가 상황에 적절치 못하여 비웃음을 사거나 의사소통이 제대로 이루어지지 못할 것이라는 데서 오는 불안감, 셋째, 학습 결과인 평가에 대한 불안감이다. 자신감이 언어 능력을 향상했는지, 언어 능력이 향상되어 자신감이 생겼는지는 닭과 달걀의 문제와 같은 것이지만 자신감과 언어 능력 향상은 상호작용하고 있음을 보여준다(Brown, 2010).

태도는 지속적 학습이나 직간접적인 경험을 통하여 형성 또는 변화될 수 있다. 외국어 학습자의 경우 학습자 자신, 목표언어나 목표권 문화, 그리고 교사나 동료에 대해 어떤 태도를 지니고 있는가가 외국어 학습에 큰 영향을 미친다. Oller(1981)는 태도와 학습은 서로 영향을 미치는 관계이므로 긍정적인 태도는 학습을 촉진하고, 좋은 학습 결과는 자신감 있는 태도 형성에 영향을 미친다고 하였다. 교사는 학생들이 의사소통할 때 실수를 두려워하지 않고, 또 교사나 동료의 시선을 의식하지 않고 목표 언어를 사용할 수 있도록 적극적인 참여를 유도하여야 한다(김정렬, 2011). 학습자가 학습 결과에 대해 만족감을 느끼는 것이 중요하다. 학습자가 학습 결과에 대해 만족한다면 학습 동기는 지속될 것이고 이는 궁극적으로 좋은 학업 성취를 가져올 것이다(하명정, 2021). 이와 같이 정의적 영역의 요인들을 간단히 살펴보았다.

3. 메타 분석

메타 분석은 어떤 주제에 관해 연구한 여러 논문의 결과를 표준화된 척도인 효과 크기로 계산하여 결과들을 종합 분석하는 방법이다. 이를 위하여 Cooper(1982)의 5단계인, 연구 주제 정하기(problem formulation), 자료 모으기(data collection), 자료 점검하기(data evaluation), 자료 분석 및 해석하기(data analysis and interpretation), 출판 및 발표하기(public presentation) 순서로 진행한다. 구체적으로, 제일 먼저 관심 있는 연구 주제를 정한다. 둘째, 정한 연구 주제에 맞는 자료를 모아야 하는데, 반복해서, 모든 관련된 자료를 사용하여 철저하게, 가능한 한 편향되지 않게, 그리고 찾고자 하는 주제와 관련된 주요 용어를 검색어로 사용하여 주제를 대표하는 연구논문을 모아야 한다(Lefebvre 외, 2022).

셋째, 수집된 연구논문을 점검할 때는 해당 연구논문을 메타 분석에 포함할 것인지 제외할 것인지를 결정하는 포함-제외 선정 기준(inclusion/exclusion criteria)을 마련한다(McKenzie 외, 2023). 포함-제외 선정 기준에서 가장 중요한 것은 효과 크기와 그 분산을 계산할 수 있는 충분한 데이터가 있는지를 점검하는 것이다. 효과 크기는 효과의 양 또는 정도를 이르는 말인데(박옥희, 윤지여, 2019), 집단의 평균 차이를 나타내는 효과 크기를 d 효과 크기(effect size) 라고 한다. 학교 현장에서 d 효과 크기를 알아보기 위해 사용되는 방법은 크게 두 가지이다(Lipsey & Wilson, 2000). 하나는, 실험 집단과 통제 집단을 이용한 실험 연구의 결과로 제시하는 독립 표본 t 검정(또는 F 검정)과 사전검사와 사후 검사 결과에서 나온 대응 표본 t 검정이 있다. 이 두 가지 형태의 평균 차이를 모아 표준화된 척도인 효과 크기로 나타내어 모집단(population)의 효과 크기를 추정할 수 있다. 선정 기준에 의해 수집된 연구논문은 코딩 매뉴얼에 맞추어 두 명 이상의 코딩자가 각각 코딩한다.

넷째, 자료 분석 및 해석 단계에서는 연구논문의 결과를 통합하여 통계 프로그램으로 메타 분석의 결과를 얻은 다음 해석한다. 효과 크기를 추정하기 위해서는 분석에 사용된 모든 연구[1]가 하나의 공통 효과 크기(common effect size)를 가지는지, 평균 효과 크기(mean effect size)를 가졌는지를 알아보아야 한다(Borenstein 외, 2007). 메타 분석에서 사용하는 모델은 크게 고정 효과 모델(fixed-effects model)과 랜덤 효과 모델(random-effects model)이다. 두 모델은 효과 크기의 분산에 따른 가중치(weights)에 따라 달라진다. 고정 효과 모델에서는 연구내 분산만을 이용하여 가중치를 계산하여 사용하고, 랜덤 효과 모델에서는

[1] 본 연구에서는 학회지나 학회 발표문에 실린 연구논문은 '연구논문(paper)'이라 칭하고, 그 연구논문에서 구체적인 효과 크기를 산출해 낸 각 연구를 '연구(study)'라 칭한다. 이것은 이전 연구(박옥희, 윤지여, 2019)에서 연구논문과 연구사례라고 칭한 것과는 다른 표기이다. 왜냐하면, 메타 분석에 자주 등장하는 'within-study variance and between-study variance'를 의미가 통하면서 좀 더 간략하고 쉽게 '연구내 분산과 연구간 분산'으로 나타내기 위해서다.

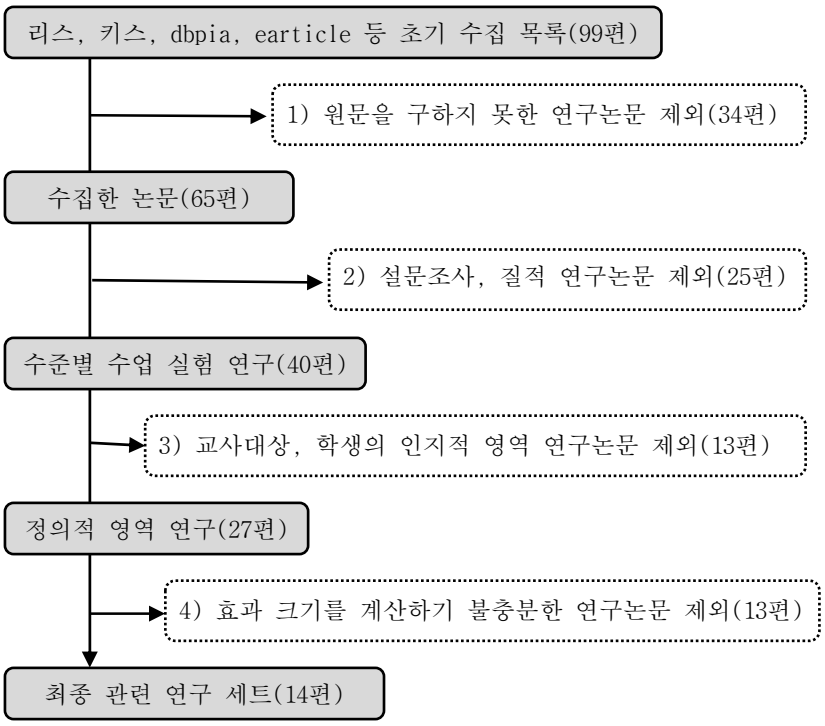
연구내 분산과 연구간 분산을 합한 분산을 이용하여 가중치를 계산한다 (Schmidt 외, 2009). 그래서 고정 효과 모델과 랜덤 효과 모델에서 사용한 개개 연구의 효과 크기뿐만 아니라 평균 효과 크기도 달라지는 것이다. 그리고 메타 분석 연구의 마지막 단계는 연구논문 게재나 연구논문을 발표하는 것이다 (Cooper, 1982).

Ⅲ. 연구 방법

1. 메타 분석 대상 연구의 선정

수준별 영어 수업에 관한 관련 연구 자료를 수집하기 위하여 *수준별 영어 수업 정의적*, *수준별 영어 수업*, *수준별 영어* 등의 키워드를 사용하였다.

그림 1
연구논문 선정 절차



그리고 이 키워드로 한국교육학술정보원의 RISS 학술 연구 정보 서비스, 한국학술정보(KISS), 누리미디어의 디비피아, 학술논문 검색서비스인 e-article 등

의 데이터베이스에서 연구논문을 검색하였다. 데이터베이스에서 논문의 제목과 초록을 살펴본 뒤, 작성한 초기 수집 목록은 99편이었다. 초기 수집 목록은 포함-제외 기준에 따라 그림 1과 같은 절차로 연구논문을 선정하였다.

첫 번째, 논문의 원문을 구하지 못한 34편을 제외하고 국내에서 출판된 원문을 구할 수 있는 연구논문들만 선정하였다. 두 번째, 65편의 연구논문 중에서 질적 연구, 설문 조사를 통한 빈도 분석 연구 등 양적 연구이지만 메타 분석에 사용할 수 없는 연구논문 25편을 제외하였다. 세 번째, 수준별 수업 실험 연구 40편의 연구논문 중에서 영어 과목의 수준별 수업에 관한 실험 연구이지만 학업 성취도 결과만 수록한 연구논문 13편을 제외하였다. 그리고 마지막으로 효과 크기와 분산을 계산할 수 있는 결과 통계치를 수록하지 않은 연구논문 13편을 제외하였다. 이러한 과정을 통해 메타 분석 연구 대상으로 최종 14편의 논문이 선정되었다. 따라서 수준별 영어 수업을 통해 변화된 정의적 영역의 평균 차이를 나타내는 d 효과 크기와 분산을 계산할 수 있는 총 5,777명을 대상으로 한 53개의 연구로 메타 분석하였다.

2. 코딩 절차와 데이터 정리

메타 분석의 코딩과 데이터 정리를 위하여, 두 명이 코딩을 맡아 개별적으로 연구논문의 내용적 측면과 효과 크기 특성 측면에 관하여 코딩하였다. 연구논문에 관한 기본적인 사항인 저자, 제목, 출판 연도, 출판 상태, 저널명과 더불어 논문의 내용 측면에서는 연구 대상의 학교와 학년, 수준별 수업 기간 및 수준별 수업 형태를 코딩하였다. 그리고 효과 크기 측면에서는 정의적 영역의 세부 항목, 효과 크기 산출 방식, 실험 통제 집단의 사례 수와 사후 평가의 평균과 표준 편차, 사전 사후 평균 차이, 편차 차이, t 값, p 값과 더불어 연구에서 사용한 데이터 추출 방법과 데이터가 있는 페이지나 표의 번호를 함께 코딩하였다. 이러한 과정을 통한 두 코딩자의 코딩 신뢰도는 표 2와 같다.

표 2
코딩 신뢰도

내용 측면 변인	%	효과 크기 측면 변인	%
학교	100	효과 크기 산출 방식	92.9
학년	85.7	데이터 페이지와 표 번호	92.9
수준별 수업 기간	100		
수준별 수업 형태	77.3		
정의적 영역 세부항목	85.7		

먼저, 연구논문에 관한 기본적인 사항인 저자, 제목, 출판 연도, 출판 상태, 저널명에서는 코딩 신뢰도가 100%였다. 그 외 수준별 수업 형태가 77.3%, 학년과 정의적 영역 세부 항목이 각각 85.7%, 효과 크기 산출 방식과 데이터 페이지와 표 번호가 92.9%, 그리고 학교와 수준별 수업 기간이 100%의 일치도를

보였다. 이후 일치하지 않는 항목들은 두 코딩자가 완전히 합의할 때까지 의논한 뒤, 14편의 연구논문의 코딩을 완료하였다. 총 14편의 연구논문에서 53개의 연구를 도출하였으며 메타 분석에 쓰일 *d* 효과 크기와 분산을 계산하였다.

표 3
조절변인과 코딩 값의 사례수

변인	학교	학년	정의적 영역	수준별 수업 형태	수업 기간
하위범주 (사례수)	초등학교(24) 중학교(8) 고등학교(6) 대학교(15)	초4(9) 초5(3) 초6(12) 중1(8) 고1(6) 대1(15)	동기(6) 만족도(9) 불안감(4) 자신감(13) 참여도(2) 태도(5) 흥미도(14)	심화보충 활동(13) 이동 수업(22) 자료 활용(18)	10주(7) 12주(10) 14주(6) 15주(3) 16주(22) 32주(1) 40주(1)
전체	53	53	53	53	50[2]

표 3에서 보듯이, 조절변인은 '학교'와 '학년', '정의적 영역', '수준별 수업 형태', '수준별 수업 기간'이다. 선정된 연구논문을 자세히 살펴보면, 14편의 연구논문은 2003년부터 2018년까지 출판되었으며 모두 학술지에 게재되었다. 연구논문에서 나온 '정의적 영역'은 동기, 만족도, 불안감, 자신감, 참여도, 태도와 더불어 흥미도가 있었다. '수준별 수업의 형태'는 심화 보충 활동 수업과 수준별 이동 수업, 그리고 수준별 자료 활용 수업으로 분류하였다. 심화 보충 활동에는 원어민 활용, 내용 중심 수업, 소집단 구성, 웹 기반 활동, 영어 교과서 내용의 재구성 등으로 수준별 활동을 구안한 수업들이었다. 수준별 자료 활용에는 사이버 가정 학습이나 딕토글로스, 영어 동화 등 다양한 자료를 수준별로 제시하는 수업들이었다. 그리고 '수준별 수업 기간'은 10주에서 40주였다.

3. 데이터 분석 방법

메타 분석 소프트웨어인 Comprehensive Meta-Analysis 4(CMA: Borenstein 외, 2022)를 이용하여 효과 크기와 분산의 계산, 그리고 메타 분석을 하였다. 실험통제 집단이 있는 독립표본에서는 수준별 수업을 하고 난 뒤에 치른 두 집단의 사후 평가의 평균과 표준 편차, 그리고 두 집단의 사례 수를 이용하여 효과 크기와 분산을 계산하였다. 한편, 사전 사후 평균 차이는 사전 평가와 사후 평가 각각의 평균과 표준 편차, 그리고 사례 수를 이용하여 계산하였다. 수집된 연구논문에서는 상관관계 계수를 제시하지 않았기 때문에 중간 정도의 상관관계가 있는 .5의 상관계수를 이용하였다.

결과 측정치는 수준별 영어 수업에 대한 정의적 영역을 검사한 실험 통제 집

[2] 3개의 연구사례는 수업 기간을 명시하지 않았음.

단의 평균 차이 25개와 사전 사후 평가의 평균 차이 28개로 구성된 g 효과 크기다. CMA에서는 Cohen의 d를 편향 표준화 평균 차이(biased standardized mean differences)로, Hedge의 g를 불편향 표준화 평균 차이(unbiased standardized mean differences)로 언급하고 있다(Borenstein 외, 2022). 그래서 연구의 불편향 표준화 평균 차이인 Hedge의 g 효과 크기와 분산을 구했다. 하지만 결과인 평균 효과 크기를 나타낼 때는 $\bar{d}$ 기호로 나타내었다. 먼저, 적절한 모델을 결정하기 위하여 연구의 이질성 검사(tests of heterogeneity 또는 Q tests)를 하였다. 그리고 선택한 모델에서 평균 효과 크기를 추정하고, 조절변인인 '정의적 영역', '학교', '학년', '수준별 수업 형태', '수준별 수업 기간'에 따라 평균 효과 크기를 추정하였다. 이때 혼합 효과 모형(mixed-effects model)을 이용한 하위 집단 분석(subgroup analysis)을 하였다. 마지막으로 출판 편향 검사를 하였다.

Ⅳ. 연구 결과

1. 이질성 검사 결과

메타 분석을 위해 수집된 모든 연구가 공통된 하나의 모집단에서 나온 것인지, 아니면 개개의 연구들이 각자의 모집단으로부터 나온 것인지 알아보기 위해 이질성 검사를 했다.

표 4
이질성 검사 결과

Q	Df	p	Birge's ratio	I^2
349.56	52	< .001	6.72	85.1%

이질성을 나타내는 지수에는 Q 값과 Birge 비율(Birge's ratio), I^2 값 등이 있는데, Higgins 외(2003)가 내놓은 기준에 의하면 "I^2 값이 25% 이하이면 이질성이 낮고, 75% 이상이면 아주 이질적이다"(박옥희, 윤지여, 2019, p. 117)라고 하였다. 표 4에서 보듯이, I^2 값이 85.1%로 아주 이질적인 집합체라고 할 수 있다. 이것은 수준별 영어 수업이 정의적 영역에 미친 영향을 나타낸 효과 크기가 하나의 공통된 모집단을 공유하는 것이 아니라는 것을 보여준다. 이는 통계적으로 유의한 이질성 검사($Q(52) = 349.56, p < .001$) 결과와 Birge 비율(6.72), I^2 값(85.1%)이 일관되게, 연구간 차이가 있다는 것을 보여준다. 따라서 랜덤 효과 모델을 이용한 랜덤 효과 평균(random-effects mean)으로 수준별 영어 수업의 정의적 영역에 관한 효과를 추정하였다.

2. 랜덤 효과 모델 결과

표 5의 랜덤 효과 모델 결과에는 평균 효과 크기 추정치(estimate), 표준 오차(standard error), z 값, 95% 신뢰 구간의 상한선과 하한선, p값을 보여준다. 추정된 모집단의 정의적 영역에 영향을 주는 수준별 영어 수업의 평균 효과 크기는 0.33이고 추정된 평균 효과 크기의 평균 오차는 0.065이며, 통계적으로 유의한 결과를 보였다($p < .001$). Cohen(1988)의 기준인 작은 효과 크기의 기준치인 0.2보다는 크고, 중간 효과 크기인 0.5보다는 작아 수준별 영어 수업은 정의적 영역에 다소 효과가 있다고 할 수 있다.

표 5
랜덤 효과 모델 결과

모델	추정치	표준 오차	z	신뢰구간		p
				하한값	상한값	
랜덤 효과	0.33	0.065	5.09	0.20	0.46	< .001

하지만 앞서 말한 85%의 연구간 차이가 있다는 것은 연구마다 모집단이 다르다는 것을 나타내므로 서로 다른 하위 영역에서 서로 다른 효과 크기를 가지고 있을 가능성이 있다. 그래서 조절변인을 이용하여 조절변인에 따라 수준별 영어 수업의 효과 크기가 어떠한지를 하위 집단 분석으로 알아보았다.

3. 하위 집단 분석

하위 집단 분석은 조절변인인 '정의적 영역'의 세부 항목을 비롯하여 '학교'와 '학년', '수준별 영어 수업 형태', '수준별 수업 기간'으로 수준별 영어 수업이 정의적 영역에 미치는 영향을 알아보았다. 하위 집단 분석은 랜덤 효과 모델과 고정 효과 모델의 형식을 혼용한 혼합 효과 모델(mixed-effects model)로 분석하였다(Borenstein, 2019). 혼합 효과 모델 결과에서 하위 집단별로 나오는 이질성 수치의 $Q_{조절변인}(df)$값은 조절변인에 대한 영가설(모든 하위 집단의 평균 효과 크기가 같다.)의 기각과 채택을 위해 확인하였고, 평균 효과 크기는 효과 크기에 대한 영가설(특정 하위 집단의 평균 효과 크기가 0이다)의 검증에 제시되는 z 값과 p 값으로 확인하였다.

1) '정의적 영역' 조절변인

결과측정치는 g효과 크기이고, 조절변인은 범주형 변인으로 하위 집단 범주는 동기, 만족도, 불안감, 자신감, 참여도, 태도, 흥미도로 일곱 가지이다. 먼저, '정의적 영역'의 하위 집단별 평균 효과 크기가 모두 같은지를 알아본 결과, 표 6과 같이 '정의적 영역' 조절변인은 통계적으로 유의하였다 ($Q_{정의적\ 영역}(6) = 41.8$,

$p < .001$). 즉, '정의적 영역' 하위 집단은 수준별 영어 수업의 효과 크기에 중요한 예측 변인이라고 할 수 있다. 또한, 표 6에 나온 하위 집단별 평균 효과 크기 추정치는 자신감과 흥미도($\bar{d}_{자신감}$[3] $= 0.68$, $\bar{d}_{흥미도} = 0.57$, 모두 $p < .001$), 그리고 참여도($\bar{d}_{참여도} = 0.51$, $p = .049$)가 유의하였다. 즉, 수준별 영어 수업은 자신감과 참여도, 흥미도에서 중간 이상의 효과가 있었다.

표 6
'정의적 영역' 조절변인의 혼합 효과 모델 결과

범주	연구사례 수	평균 효과 크기 추정치	표준 오차	z	p
동기	6	0.17	0.157	1.08	.280
만족도	9	-0.12	0.107	-1.09	.274
불안감	4	-0.05	0.168	-0.30	.763
자신감	13	0.68	0.105	6.41	< .001
참여도	2	0.51	0.257	1.97	.049
태도	5	0.12	0.166	0.74	.46
흥미도	14	0.57	0.100	5.65	< .001
$Q_{정의적\ 영역}(6) = 41.8, p < .001$					

2) '학교' 조절변인

'학교' 조절변인의 하위 집단은 초등학교, 중학교, 고등학교, 대학교로 네 개이고 결과측정치는 g효과 크기이다.

표 7
'학교' 조절변인의 혼합 효과 모델 결과

범주	연구사례 수	평균 효과 크기 추정치	표준 오차	z	p
초등학교	24	0.67	0.076	8.82	< .001
중학교	8	0.07	0.107	0.69	.488
고등학교	6	0.65	0.118	5.50	< .001
대학교	15	-0.12	0.076	-1.54	.124
$Q_{학교}(3) = 67.03, p < .001$					

학교별로 평균 효과 크기가 서로 같은지를 알아본 결과, 표 7에서 보듯이, '학교' 조절변인은 통계적으로 유의하였다($Q_{학교}(3) = 67.03, p < .001$). 즉, '학교' 조절변인이 수준별 영어 수업의 효과 크기에 중요한 예측 변인이라는 것을 알 수 있다. 또한, 표 7의 학교별 평균 효과 크기 추정치는 초등학교와 고등학교만 ($\bar{d}_{초} = 0.67$, $\bar{d}_{고} = 0.65$, 모두 $p < .001$) 유의하였다. 즉, 수준별 영어 수업은 초등학교와 고등학교에서 중간 이상의 효과가 있었다.

[3] $\bar{d}_{자신감}$ 는 자신감 그룹의 g 효과 크기의 평균 추정치를 의미한다.

3) ‘학년’ 조절변인

‘학년’ 조절변인은 초등학교 4학년, 5학년, 6학년, 중학교 1학년, 고등학교 1학년, 대학교 1학년으로 여섯 개의 하위 집단으로 구성된다. 표 8에서 보듯이, ‘학년’ 조절변인도 통계적으로 유의하여($Q_{학년}(5) = 83.81, p < .001$), ‘학년’ 변인도 수준별 영어 수업의 효과 크기에 중요한 예측 변인이라고 할 수 있다. ‘학교’와 마찬가지로 표 8에 나온 학년별 평균 효과 크기 추정치는 초등학교 4, 5, 6학년과 고등학교 1학년만 ($\bar{d}_{초4} = 0.82$, $\bar{d}_{초5} = 1.07$, $\bar{d}_{초6} = 0.47$, $\bar{d}_{고1} = 0.65$, 모두 $p < .001$) 유의하였다. 초등학교 4학년과 5학년에서는 매우 큰 효과가 있었고 (Sawilowsky, 2009), 초등학교 6학년과 고등학교 1학년에서는 중간 정도의 효과가 있었다.

표 8
‘학년’ 조절변인의 혼합 효과 모델 결과

범주	연구사례 수	평균 효과 크기 추정치	표준 오차	z	p
초등 4학년	9	0.82	0.115	1.04	< .001
초등 5학년	3	1.07	0.231	1.53	< .001
초등 6학년	12	0.47	0.102	0.67	< .001
중학교 1학년	8	0.07	0.100	-0.13	.476
고등학교 1학년	6	0.65	0.111	0.43	< .001
대학교 1학년	15	-0.12	0.071	-0.26	.099
$Q_{학년}(5) = 83.81, p < .001$					

4) ‘수준별 수업 형태’ 조절변인

표 9
‘수준별 수업 형태’ 조절변인의 혼합 효과 모델 결과

범주	연구사례 수	평균 효과 크기 추정치	표준 오차	z	p
수준별 이동 수업	22	0.15	0.091	1.67	.095
심화보충 활동	13	0.76	0.106	5.62	.005
수준별 자료 활용	18	0.30	0.135	2.79	< .001
$Q_{수준별\ 수업\ 형태}(2) = 14.07, p = .001$					

‘수준별 수업 형태’ 조절변인의 범주는 수준별 이동 수업, 심화 보충 활동, 수준별 자료 활용 형태로 세 가지이다. 표 9에서 보듯이, ‘수준별 수업 형태’ 조절변인 역시 통계적으로 유의하여($Q_{수준별\ 수업\ 형태}(2) = 14.07$, $p = .001$), ‘수준별 수업 형태’도 수준별 영어 수업의 효과 크기를 예측할 수 있는 중요한 변인이라고 할 수 있다. 표 9의 수준별 수업 형태별 평균 효과 크기 추정치는 심화 보충 활동 형태와 수준별 자료 활용 형태만($\bar{d}_{심화\ 보충\ 활동} = 0.76$, $\bar{d}_{수준별\ 자료\ 활용} = 0.30$, 각각 $p = .005$, $p < .001$) 유의하였다. 즉, 심화 보충 활동 수업은 큰 효과가 있고,

수준별 자료 활용 수업은 작은 효과가 있으며, 수준별 이동 수업은 효과가 없었다.

5) '수준별 수업 기간' 조절변인

'수준별 수업 기간' 변인은 수업 기간을 명시하지 않은 세 개의 연구와 가외치(outlier)인 수업 기간 32주와 40주인 두 개의 연구를 제외하고 48개 연구로 하위 집단 분석하였다. '수준별 수업 기간' 조절변인의 범주는 10주와 12주, 14주, 15주, 16주로 다섯 가지이다. 표 10에서 보듯이, '수준별 수업 기간' 조절변인 역시 통계적으로 유의하여($Q_{수준별\ 수업\ 기간}(4) = 17.71$, $p = .001$), 수준별 영어 수업의 효과 크기에 중요한 예측 변인이라고 할 수 있다. 표 9의 수준별 수업 기간별 평균 효과 크기 추정치는 10주와 14주, 16주만($\bar{d}_{10주} = 0.72$, $\bar{d}_{14주} = 0.71$, 각각 $p < .001$, $\bar{d}_{16주} = 0.16, p = .049$) 유의하였다. 즉, 수준별 수업 기간이 10주와 14주일 때 큰 효과가 있고, 16주일 때는 아주 작은 효과가 있으며, 수업 기간 12주와 15주에는 효과가 없었다.

표 10
'수준별 수업 기간' 조절변인의 혼합 효과 모델 결과

범주	연구사례 수	평균 효과 크기 추정치	표준 오차	z	p
10주	7	0.72	0.172	4.19	< .001
12주	10	0.13	0.119	1.10	.27
14주	6	0.71	0.185	3.82	< .001
15주	3	-0.04	0.212	-0.20	.84
16주	22	0.16	0.081	1.97	.049
$Q_{수준별\ 수업\ 기간}(4) = 17.71, p = .001$					

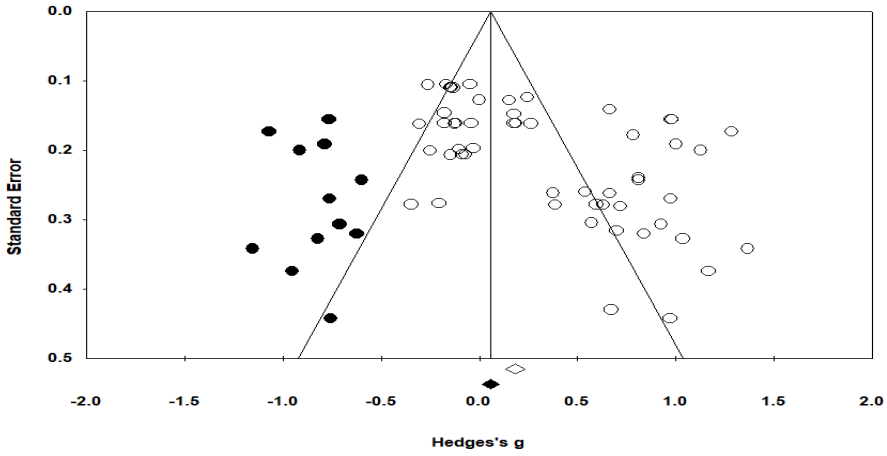
4. 출판 편향 검사(publication bias) 결과

출판 편향은 효과가 크거나 유의한 결과를 도출해 내지 못해서 세상에 나오지 못하고 숨겨진 연구논문이 있을 거라는 생각에서 비롯된 용어이다(Rosenthal, 1979). 효과가 크고 유의한 결과가 나온 연구논문들이 상대적으로 그다지 크지 않은 효과를 보이는 연구논문들보다 학술지나 학회에 게재 또는 발표되는 확률이 크다는 것이다. 그러므로 출판된 연구논문들을 모아 메타 분석한 연구는 편향이 있을 수 있다. 즉, 메타 분석을 위해 모인 연구논문들의 진 효과 크기가 과대 추정되는 경향이 있으므로 출판 편향의 정도를 평가하는 것이 중요하다 (Borenstein, 2019).

출판 편향 검사는 깔때기 도표(funnel plot)와 다듬고 채우기(trim and fill) 기법으로 알아보았다(박옥희, 윤지여, 2019). 깔때기 도표는 가로축에 g 효과 크기를, 세로축에 표준 오차를 두고 연구를 점으로 나타내었다. 보통 사례수가 많은 연

구(표준 오차가 적은 연구)는 도표의 위쪽에, 사례수가 적은 연구(표준 오차가 큰 연구)는 아래쪽에 점이 찍힌다. 그리고 효과 크기가 큰 연구는 오른쪽에, 효과 크기가 작은 연구는 왼쪽에 점이 찍히는데, 평균을 중심으로 좌우 대칭을 이루면 출판 편향이 존재하지 않는다고 한다.

그림 2
깔때기 도표



그러나 그림 2에서 보듯이, 흰색 점은 분석에 사용된 연구의 효과 크기와 표준 오차에 맞춰 찍힌 것으로 출판 편향이 존재한다고 할 수 있다. 즉, 본 메타 분석에는 효과 크기가 작은 연구들이 부족하다는 것을 말해준다. 또한, 그림 2의 검은 점은 결여된 연구(missing studies)를 표시한 것으로, 다듬고 채우기 기법에 따르면 도표의 왼쪽 부분에 13개의 연구가 결여되었다고 나왔다. 만약에 13개의 연구를 포함한 66개의 연구로 메타 분석한다면 랜덤 효과 평균이 0.33에서 0.13로 수정되리라 예측하였다. 따라서 본 메타 분석은 출판 편향이 있으며 랜덤 효과 평균인 0.33은 과대 추정되었을 가능성이 있다고 할 수 있다.

하지만, Borenstein(2019)은 메타 분석을 하면서 흔히 저지를 수 있는 실수 중의 하나가 출판 편향 검사의 해석이라고 하면서, 출판 편향 검사 결과의 올바른 해석을 언급하였다. 다시 말해, 결과의 타당성을 생각할 때 출판 편향의 잠재적 영향력을 고려할 필요는 있지만, 출판 편향이 있다고 해서 자동으로 메타 분석 결과가 틀렸다고 할 수 없다. 따라서 본 메타 분석은 출판 편향이 존재하여 평균 효과 크기를 과대 추정했을 수는 있지만, 연구 결과가 타당하지 않다고는 단정할 수 없다는 것이다.

Ⅴ. 결론 및 제언

수준별 영어 수업에서 인지적 영역의 효과를 알아본 박옥희와 윤지여(2019)의 후속 연구로, 수준별 수업의 정의적 영역에 대한 효과를 메타 분석을 통해 알아보았다. 이를 위해 필요한 데이터가 충분하다고 판별된 최종 14편의 연구 논문을 통해 총 55개의 개별 연구를 추출하였다. 2003년부터 2018년 사이에 발표된 논문에서 총 5,777명을 대상으로 하여 정의적 영역, 수준별 영어 수업 형태 및 수업 기간, 학교와 학년별로 평균 효과 크기를 살펴본 결과는 다음과 같다.

첫 번째 연구 문제인 수준별 영어 수업이 정의적 영역에 미치는 평균 효과 크기는 0.33이었고, 이는 중간 정도의 효과라고 볼 수 있다. 인지적 영역에서의 평균 효과 크기 0.74 (박옥희, 윤지여, 2019)보다는 작았다. 따라서 수준별 수업은 인지적 영역에서는 효과가 크고, 정의적 영역에서는 효과가 보통이라고 할 수 있다.

두 번째 연구 문제인 조절변인별 효과 크기의 양상은 다음과 같다. 먼저 첫째, 정의적 영역 조절변인인 흥미도, 자신감, 참여도, 만족도, 불안감, 동기, 태도는 통계적으로 유의하였고($p < .001$), 자신감과 흥미도($\bar{d}_{자신감} = 0.68$, $\bar{d}_{흥미도} = 0.57$, 모두 $p < .001$), 참여도($\bar{d}_{차여도} = 0.51$, $p = .049$)에서 중간 이상의 효과를 보였다. 선행연구에서는 학습자의 수준에 따라 만족도에 차이가 있다고 보고되었지만(김윤정, 2015; 김영숙, 이지연, 2009; 박지영 외, 2012), 본 연구에서는 수준별 영어 수업에 관한 만족도는 효과가 없는 것으로 나타났다. 반면, 흥미도와 자신감에서는 선행연구(안애랑, 박은수, 2014)에서처럼 효과가 있었다.

둘째, 학교와 학년 별로 살펴본 결과, 학교와 학년 조절변인 모두 통계적으로 유의하였고($p < .001$), 초등학교와 고등학교($\bar{d}_{초} = 0.67$, $\bar{d}_{고} = 0.65$, 모두 $p < .001$)에서 중간 이상의 효과가 있었다. 학년별로는 초등학교 4학년, 5학년, 6학년과 고등학교 1학년에서 ($\bar{d}_{초4} = 0.82$, $\bar{d}_{초5} = 1.07$, $\bar{d}_{초6} = 0.47$, $\bar{d}_{고1} = 0.65$, 모두 $p < .001$) 통계적으로 유의하였다. 초등학생 4학년과 5학년에서는 큰 효과가 있었고, 초등학교 6학년과 고등학교 1학년에서는 중간 정도의 효과가 있었다. 인지적 영역에서의 학교와 학년별 조절변인도 통계적으로 유의하다는 선행연구(박옥희, 윤지여, 2019)결과가 있었고, 구체적으로 초등학교, 고등학교, 대학교에서 시행했을 때 효과가 크게 나타났었다. 그리고 학년별로는 대학교 1학년부터 4학년, 고등학교 1학년, 초등학교 3학년부터 6학년에게 효과가 좋은 것으로 보고되었다. 따라서 수준별 수업은 초등학생과 고등학생들의 인지적 영역과 정의적 영역에 모두 중간 이상의 효과가 있고, 학년별로는 초등학교 4, 5, 6학년과 고등학교 1학년생들에게 특히 효과가 좋은 것을 알 수 있었다.

셋째, 수준별 수업 형태 변인에서는 심화 보충 활동 형태와 수준별 자료 활용 형태만($\bar{d}_{심화\ 보충\ 활동} = 0.76$, $\bar{d}_{수준별\ 자료\ 활용} = 0.30$, 각각 $p = .005, p < .001$) 통계적으로 유의하였다. 심화 보충 활동 수업에서는 정의적 영역에 큰 효과가 있고,

수준별 자료 활용 수업에서는 중간 정도의 효과가 있었다. 심화 보충 수업에서의 큰 효과에 대해 신재한(2007)은 학습자의 능력, 수준, 적성 등을 고려하여 단위 수업 시간의 결손을 최소화함으로써 학습자들의 관심과 흥미, 자신감을 고양할 수 있다고 하였다

마지막으로 수준별 수업의 기간은 수업 기간 10주와 14주만($\bar{d}_{10주}$ = 0.72, $\bar{d}_{14주}$ = 0.71, $p < .001$) 통계적으로 유의하였다. 즉, 수준별 수업 기간이 10주와 14주일 때 큰 효과가 있고, 16주일 때는 아주 작은 효과가 있으며, 수업 기간 12주와 15주에는 효과가 없었다.

본 연구의 제한점으로 표준오차를 조정하는 강한 분산 추정 기법(robust variance estimation: RVE)을 사용하지 못한 것을 들 수 있다. 본 메타 분석에서 하나의 연구논문에서 최대 7개의 연구가 있었다. 이렇게 하나의 연구논문에서 다수의 연구가 나올 때 이 연구들은 서로 관련이 있어서, 효과 크기의 분산을 과대 또는 과소 추정할 수 있는 우려가 있기 때문이다(Scammacca 외, 2014). 그러나 수준별 수업에서의 인지적 영역에 대한 효과를 조사한 박옥희와 윤지여(2019)의 선행연구에 이어 정의적 영역에서의 효과를 살펴본 것에서 의미를 찾을 수 있다. 이러한 연구 결과는 앞으로 수준별 수업의 교과과정을 수립하고 개선하는 데 이바지하는 바가 있으리라 기대한다.

참고문헌

*표시는 메타 분석 대상 논문임.

권선희, (2012). 대학 교양 영어 프로그램 개선 방안 연구: 교육과정 구성 및 학습자 인식을 중심으로. *영어어문교육, 18*(4), 211-235.

김영숙, 이지연. (2009). 대학 교양 영어 수준별 수업에 대한 학습자와 교수자의 인식에 관한 연구. *영어교육, 64*(4), 337-368.

김영현. (2002). 초등학교 영어 과 보충, 심화형 수준별 수업 지도 방안과 실제. *초등영어교육, 8*(2), 63-96.

김윤정. (2015). 대학의 수준별 영어 수업에 대한 연구. *영어교육연구, 27*(3), 7-103.

*김은주, 나경희. (2015). 수준별 영어 수업 실시 여부에 따른 중학생의 영어능력 향상도 분석. *영어영문학 연구, 57*(1), 132-156.

김은주, 성귀복. (2016). 형태초점 기법 활용 수업이 초등영어 성취도에 미치는 영향. *교육문화연구, 22*(6), 399-423.

김정렬. (2011). 영어과 한국인 교사와 원어민 영어 보조교사 협력 수업의 효과 메타 분석. *영어교과교육, 10*(3), 91-113.

*김정렬, 이선아, 김윤정. (2005). 초등학교 영어, 과학 내용 중심 통합 수준별 수업의 효과. *영어교육연구, 17*(2), 105-134.

*김정렬, 황선영. (2007). 초등학교 영어 사이버 가정학습 수준별 콘텐츠를 통한 교실수업 연계방안. *Multimedia Assisted Language Learning, 10*(1), 130-154.

*김현옥, 이혜경. (2010). 대학 교양 영어 수준별 프로그램 운영방안. *영어교육연구, 22*(4),75-101.

*나기연, 정미란. (2003). 초등영어교육에서 영어 원어민을 활용한 수준별 수업의 효과. *영어교육연구, 8*(1), 131-152.

남지영. (2020). 이공계대학 교양 영어의 수준별 수업 운영과 선호도에 관한 연구. *외국어교육, 27*(4), 73-96.

박상옥. (2004). 효과적인 중학교 수준별 영어 수업 방안에 대한 연구. *한국 교원교육연구, 21*(3), 71-89.

*박선호, 송현진. (2013). 딕토글로스 활용 수준별지도가 초등학생 영어능력신장 및 정의적 영역에 미치는 효과. *응용언어학, 29*(4), 249-288.

박옥희. (2012). 대학 영어 캠프 참가대학생들의 또래 ESL튜터와 성인 원어민 교사에 대한 정의적 영역 비교 교찰. *교양교육연구, 6*(2), 423-452.

박옥희, 윤지여. (2019). 메타 분석을 통한 수준별 영어 수업의 효과 크기 분석: 언어의 네 가지 기능을 중심으로. *영어어문교육, 25*(4), 107-130.

박지영, 오준일, 이인규. (2012). 대학의 수준별 토익 수업에 대한 학습자의 만족도에 관한 연구. *영어교육연구, 24*(3), 257-276.

서정아. (2018). 대학 교양 영어 수준별 수업에 대한 학생들의 인식에 관한 연구. *교육문화연구, 24*(5), 355-374.

*신재한. (2007). 심화보충형 수준별 영어수업 모형 개발 적용이 말하기 듣기 능력과 자기주도적 학습태도에 미치는 효과. *교육과정 평가, 10*(2), 103-124.

심지현, 임현우. (2017). 학습자 변인 및 분반 형태에 따른 영어 과 수준별 수업의 효과. *중등영어교육, 65*(4), 883-912.

안애랑, 박은수. (2014). 수준별 영어수업이 중학교 하반 학생들의 정의적 영역에 미치는 영향. *학습자중심교과교육연구, 14*(9), 47-64.

양정임. (2019). 학습자 수준에 따른 교양 영어과 수준별 수업의 효과. *중등영어교육, 12*(3), 221-235.

*오희정, 이미경. (2015). 수준별 학습자의 정의적 요인이 영어 학습에 미치는 영향. *영어교육연구, 27*(4), 265-286.

이경화, 김경혜, 한문섭. (2016). 영어 수준별 이동수업이 초등학교 학생들의 정의적 영역에 미치는 영향 및 개선 방안. *학습자중심교과교육연구, 16*(2), 667-692.

이지은, 장종득. (2014). 수준별 영어 수업이 학업 성취에 미치는 영향. *언어과학연구, 71*, 197-216.

임정완. (2014). 영어과 수준별 이동수업에 대한고등학교 학생과 교사의 인식 연구. *영어교과교육, 13*(1), 111-133.

*전영주, 강용구, 정해성. (2003). WBI 수준별 수업을 통한 영어 의사소통능력 향상 방안. *영어어문교육, 12*(4), 117-138.

*정말선, 우길주. (2006). 수준별 소집단 학습이 초등영어 읽기 능력 및 태도에 미치는 영향. *Studies in English Education, 11*(2), 85-103.

*정행, 노영신. (2009). 수준별 과업 중심의 수업을 통한 초등영어 읽기 쓰기 지도와 효과 분석. *영어영문학21, 22*(4), 157-178.

주미란. (2021). 대학 영어의 수준별 수업을 위한 IPA분석. *교양교육연구, 15*(2), 11-23.

*하명애. (2018). 원어민교수의 교양 영어 수준별 수업에 대한 학습자와 교수자의 인식 연구. *현대영어영문학, 62*(1), 223-250.

하명정. (2021). 온라인 PBL 수업에서 학습자 만족도와 학습효과를 예측하는 요인분석. *학습자중심교과교육연구, 21*(24), 149-165.

하명호. (2014). 수준별 토익 집중 프로그램 운영 효과. *영어영문학 연구, 56*(4), 609-630.

*하수지, 박기화. (2017). 어휘 학습전략을 활용한 영어 동화 수준별 수업이 초등학생의 영어 읽기 능력에 미치는 영향. *초등영어교육, 23*(2), 143-169.

*홍은화, 김경자. (2014). 수준별 영어 이동수업에서 고등학생들의 영어 성취도, 흥미도, 자신감의 변화. *외국어교육연구, 28*(2), 233-254.

Borenstein, M. (2019). *Common mistakes in meta-analysis and how to avoid them*. Biostat.

Borenstein, M., Hedges, L. E., Higgins, J. P. T., & Rothstein, H. R. (2022). *Comprehensive Meta-Analysis* (Version 4) [Software]. https://Meta-Analysis.com

Borenstein, M., Hedges, L. V., & Rothstein, H. R. (2007). *Meta-analysis fixed effect vs. random effects*. https://www.meta-analysis.com/downloads/M-a_f_e_v_r_e_sv.pdf

Brown, H. D. (2007). *Teaching by principles: An interactive approach to language pedagogy* (2nd ed.). Longman.

Brown, H. D. (2010). *Principles of language learning and teaching* (5th ed.). Prentice Hall.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Cooper, H. M. (1982). Scientific guidelines for conducting integrative research reviews. *Review of Educational Research, 52,* 291-302.

Deci, E. (1975). *Intrinsic motivation*. Plenum Press.

Domina, T., Penner, A., & Penner, E. (2017). Categorical inequality: Schools as sorting machines. *Annual Review of Sociology, 43*, 311-330.

Donaldson, M. L., LeChasseur, K., & Mayer, A. (2017). Tracking instructional quality across secondary mathematics and English language arts classes. *Journal of Educational Change, 18*(2), 183-207.

Gamoran, A., & Mare, R. D. (1989). Secondary school tracking and educational inequality: Compensation, reinforcement, or neutrality? *American Journal of Sociology, 94*, 1146-1183.

Gardner, R. (1985). *Social psychology and second language learning: The role of attitude and motivation.* Edward Arnold.

Gardner, R., & Lambert, W. (1972). *Attitudes and motivation in second language learning*. Newbury House.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327*, 557-560.

Horwitz, E., Horwitz, M., & Cope, J. (1986). Foreign language classroom anxiety. *Modern Language Journal, 70*, 125-132.

Krashen, S. (1982). *Principles and practice in second language acquisition*. Pergamon Press.

Krashen, S. (1983). Newmark′s ignorance hypothesis and current second language acquisition theory. In S. Gass, & L. Selinker (Eds.), *Language transfer in language learning* (pp. 135-153). Newbury House.

Kulick, J. A. (1992). An analysis of the research on ability grouping: Historical and contemporary perspectives. *Ability Grouping*, 1-45.

Lefebvre, C., Glanville, J., Briscoe, S., Featherstone, R., Littlewood, A., Marshall, C., Metzendorf, M-I., Noel-Storr, A., Paynter. R., Rader, T., Thomas, J., & Wieland, L. S., (2022). Searching for and selecting studies. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane handbook for systematic reviews of interventions* version 6.3. Cochrane. https//training.cochrane.org/handbook

Lipsey, M. W., & Wilson, D. (2001). *Practical meta-analysis*. Sage.

McKenzie, J. E., Brennan, S. E., Ryan, R. E., Thomson, H. J., Johnston, R. V., & Thomas, J. (2023). Defining the criteria for including studies and how they will be grouped for the synthesis. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane handbook for systematic reviews of interventions* version 6.4. Cochrane. https//training.cochrane.org/handbook

Oller, J. (1981). Language as intelligence? *Language Learning, 31*, 465-492.

Oxford, R. (1999). Anxiety and the language learner: New insights. In J. Arnold (Ed.), *Affect in language learning* (pp. 58-67). Cambridge University Press.

Rogers, C. R. (1967). *The therapeutic relationship and its impact: A study of psychotherapy with schizophrenics.* Ohio University Press.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638-641.

Sawilowsky, S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods, 8*(2), 467-474.

Scammacca, N., Roberts, G., & Stuebing, K. K. (2014). Meta-analysis with complex research designs: Dealing with dependence from multiple measures and multiple group comparisons. *Rev Educ Res, 84*(3), 328-364.

Schmidt, F. L., Oh, I., & Hayes, T. L. (2009). Fixed- versus random-effects mod els in meta-analysis: Model properties and an empirical comparison of diff erences in results. *British Journal of Mathematical and Statistical Psychology, 6 2*, 97-128.

Slavin, R. E. (1987). Ability grouping and student achievement in elementary schools: A best evidence synthesis. *Review of Educational Research, 57*, 293-336.

Tomlinson, C. A. (2014). *The differentiated classroom: Responding to the needs of all learners* (2nd ed.). ASCD.

예시언어(Examples in)：English
적용가능 언어(Applicable Language)：English
적용가능 수준(Applicable Levels)：Elementary, Secondary

윤지여(교사)
경남 사천 정동초등학교
52523 경상남도 사천시 정동면 대곡 1길 25
Email: yjymodu@empas.com

박옥희(교수)
배재대학교 주시경대학 기초교육부
35345 대전광역시 서구 배재로 155-40
Email: okpark@pcu.ac.kr

Received: June 23, 2023
Reviewed: September 1, 2023
Revised version received: September 14, 2023