

메타 분석을 통한 수준별 영어 수업의 효과 크기 분석: 언어의 네 가지 기능을 중심으로

박 옥 회*
(배재대학교)
윤 지 여**
(사천 정동초등학교)

Park, OkHee & Yun, Jiyeo (2019). A meta-analysis on the effects of four skills in level differentiated English classes. *English Language & Literature Teaching* 25(4), 107-130.

Many studies have reported on the effects of level differentiated English classes, but the results are inconclusive. This study aims to investigate the overall effect of four skills in level differentiated English classes. The research is based on 11 primary studies that have been published in Korea from 2003 to 2017. The results indicate that level differentiated English classes had an overall high effect size ($\bar{d} = 0.74$). More specific findings of meta-analysis are as follows. First, grouping classes based on ability is most effective for the lower level group ($\bar{d} = 0.97$). Second, applying level differentiated English classes to students in elementary school, middle school, high school and university, the effects are positive with the exception of middle school students. Third, based on language skills, it is effective for improving listening ($\bar{d} = 0.88$), reading ($\bar{d} = 0.87$), and speaking ($\bar{d} = 0.85$). Fourth, when considering how to group students according to ability, grouping by pre-evaluation ($\bar{d} = 0.98$) is more effective than relying on the students' achievements in previous term ($\bar{d} = 0.68$). The pedagogical implications and suggestions are further discussed.

[meta-analysis/four skills/level differentiated English classes/effect size, 메타 분석/언어의 네 가지 기능/수준별 영어 수업/효과 크기]

* 제 1저자
** 교신저자

1. 서론

제 2외국어로서의 영어(EFL, English as a foreign language)를 배우는 우리나라 학습자들은 영어에 대한 능력과 적성, 흥미와 필요, 그리고 주변 여건의 차이로 인해 영어성취도에 큰 차이를 보이고 있다. 수준별 수업은 이러한 개인차를 고려하여 같은 수준의 학습자들을 모아 학습자들의 수준과 욕구에 맞게 가르치는 학습자 중심 교육이다. 그러나 수준별 교육이 지향하는 교육 철학과 달리 수준별 수업의 교육 효과에 대한 논쟁은 국내외적으로 계속적으로 제기되고 있다. 우리나라의 상황만 보더라도 획일적 교육에 대한 대안으로 제 7차 교육과정부터 수준별 수업이 공교육 현장을 중심으로 전면 시행되었다가 현 입시교육 체제에 안 맞는 비현실적인 정책이라는 비판에 부딪쳐 7차 개정 교육과정부터는 학교장의 재량에 따라 수준별 수업이 실시되고 있다(심지현 & 임현우, 2017). 수준별 수업이 오래 전부터 시행되어 오고 있는 해외에서도 수준별 수업에 대한 대립된 의견이 공존하고 있는데, 수준별 수업이 학습 효율성을 높이고 교육의 질을 높인다(Kulick 1992; Slavin, 1987)는 긍정적인 주장과 수준별 수업은 하위 학습자들에게 불리하고, 학력 격차만 키운다(Domina, Penner, & Penner, 2017; Donaldson, LeChasseur, & Mayer, 2017; Gamoran & Mare, 1989)는 부정적인 주장이 있다.

국내에서도 수준별 수업에 관한 연구가 꾸준히 이루어지고 있는데 경향을 보면 수준별 수업의 운영 방식과 실태 조사(김영현, 2002; 박상옥, 2004), 수준별 수업에 대한 인식 및 만족도 조사(김영숙 & 이지연, 2009; 박지영, 오준일, & 이인규, 2012; 박찬규, 2013; 하명애, 2018), 학습자의 정의적 영역 조사(권영은 & 박은수, 2011; 민지아 & 조윤경, 2011; 이경화, 김경혜, & 한문섭, 2016; 황여정, 2008)와 성취도 조사(김윤정, 2015; 김은주 & 성귀복, 2016; 박선호 & 송현진, 2013; 이지은 & 장종득, 2014; 하명호, 2014; 홍은화 & 김경자, 2014) 등으로 나누어 볼 수 있다. 이와 같이 수준별 수업에 대한 국내 연구는 초등학생부터 대학생을 대상으로 다양한 연구 주제로 이루어지고 있으나 수준별 수업을 경험한 학습자의 정의적 영역 및 성취도에 관한 연구 결과가 상이하게 보고되고 있다. 따라서 서로 다른 이 선행연구 결과들을 하나로 모아 분석하여 보다 객관적이고 종합적인 결론을 도출해내야 할 필요성이 있다. 이에 본 연구는 메타 분석을 통해 수준별 수업의 효과를 종합적으로 살펴보고 수준별 수업에 대한 바람직한 방향에 대해 생각해보고자 한다.

메타 분석은 어떤 연구의 통계적인 유의성보다는 효과의 정도에 관심을 두고 연구의 특성과 효과 크기(effect size)의 관계를 탐색하여 필요한 정보를 얻어내는 체계적인 문헌 연구이다(Polanin & Pigott, 2013). 즉, 기존 개별 연구물의 결과에 대한 수치적인 분석을 통해 통합적이고, 종합적이며 일반화된 결론을 도출하는 연구로 분석에 대한 분석(analysis of analysis)이라고

한다(오성삼, 2002; 이제영, 문은주 & 박옥희, 2013; 이제영 & 박옥희, 2016). 이와 같이 메타 분석은 유사한 주제의 연구를 모아 모집단의 효과 크기를 추정하는 것으로(Kromrey & Hogarty, 2002) 주로 정책 연구자들이 어떤 정책을 결정하는데 필요한 정보들을 얻기 위해 관련 연구들을 한데 모아 분석해보는 양적 연구 도구 중의 하나이다.

본 연구는 우리나라 학습자를 대상으로 수준별 영어 수업을 일정기간 시행하고 영어 능력 향상 정도를 검사한 연구들을 모아 효과 크기를 살펴보았다. 구체적으로 수준을 고려한 분반 형태(상, 중, 하), 학년(초등학생, 중학생, 고등학생, 대학생), 언어 기능(듣기, 읽기, 말하기, 쓰기), 그리고 수준을 나누는 방법(사전평가, 직전 학기 성취결과, 학생들 선택)에 따라 효과 크기가 어떠한지를 살펴보았다. 이 연구 결과는 수준별 수업의 교과과정 수립 및 개선에 도움이 될 것으로 기대한다.

구체적인 연구 문제는 다음과 같다.

연구 문제 1: 수준별 영어 수업의 효과 크기는 어떠한가?

연구 문제 2: 수준별 영어 수업의 조절 변인인 분반 형태, 학년, 언어 기능, 수준 구분 방법에 따라 수준별 영어 수업의 효과 크기는 어떠한가?

II. 이론적 배경

1. 수준별 수업

수준별 수업은 학생들의 능력이나 수준, 개인차에 따라 교육목표, 교육내용, 교육방법 등을 달리하는 학습자 중심의 교육 과정이다(Brown, 2007). Tomlinson(2014)는 수준별 수업에서 고려해야 할 내용을 다음 세 가지로 정리하였다. 첫째, ‘무엇을 차별화할 것인가(Differentiate what?)’의 문제로 교과내용, 교육과정 또는 학습 환경의 차별화를 말한다. 둘째, ‘어떻게 차별화할 것인가(Differentiate how?)’의 문제로 학습자의 특성이나 흥미, 개인 경험 등을 반영하는 것을 의미한다. 셋째는 ‘왜 차별화할 것인가(Differentiate why?)’의 문제로 학습의 수월성, 학습 동기 부여와 학습의 효율성을 고민한다(김영숙 & 이지연, 2009). 또 Slavin(1987)은 수준별 수업이 성공하려면 수준 단계를 구분할 때 학습자의 언어능력을 고려해야 하고, 학습자의 개별적인 학습능력에 따라 교과과정을 편성해서 차등적으로 학습내용을 적용해야 한다고 하였다.

그러나 실제 수준별 수업의 효과에 대해서는 찬반 의견이 합의를 보지 못하고 있는 상태이다. 수준별 수업의 효과에 대한 해외 연구 보고를 보면 수준별 수업은 학생들의 학업성취라는 인지적 측면뿐만 아니라 자아개념의 형

성이라는 정의적 측면에서도 긍정적인 효과가 있다(Ansalone, 2001)는 주장이 있다. 또한, 수준별 집단의 학습 효과를 메타 분석으로 분석한 Kulick 와 Kulick(1982)는 수준별 수업이 모든 수준에 다소의 효과가 있다고 하였다. 반면, 수준별 수업의 효과에 대한 부정적인 주장으로 먼저, 상위 집단과 하위 집단간의 성취 수준의 불평등 심화만 가속된다(Gamoran & Mare, 1989)는 주장과 하위 집단 학생들의 자존감이 낮아져서 학교에 대해 부정적인 태도를 가질 수 있다(Abraham, 1989; Ball, 1981)는 주장이 있었다. 그리고 동질 집단에서 학습이 가장 효과적이라는 수준별 수업 가설의 근거가 희박하고, 또 실증적으로도 검증되지 않았으며(Wilson & Ribovich, 1973), 그 어떤 수준의 학생도 수준별 수업으로 혜택을 보지 못한다(Oakes, 1990)라는 주장이 맞다고 있다.

국내 연구를 보면, 수준별 수업에서의 성취도 향상에 관한 연구 결과가 상이했는데, 영어 수준이 낮은 학습자들의 영어 성적이 향상되었다는 연구(강보람 & 조윤경, 2012; 김윤정, 2015; 박선호 & 송현진, 2013; 이지은 & 장종득, 2014; 정양수, 2014; 하수지 & 박기화, 2017) 결과와 함께 반대로 하위권 학생들의 성적이 낮아졌다는 연구 결과(이화자 & 정정옥, 2005)가 있고, 반면, 상위권 학생들의 영어 수준이 향상되었다(김은주 & 나경희, 2015; 김은주 & 성귀복, 2016; 정혜옥 & 박부남, 2014)는 연구 결과가 함께 보고되고 있어 선행 연구로 수준별 수업에 대해 일반화할 수 있는 결론을 이끌어내기에는 복잡한 양상을 보이고 있다.

강보람과 조윤경(2012)는 부산소재 중학교 1학년과 2학년 학생을 대상으로 사전평가를 통해 상중하 세 개 분반으로 나누어 1년간 이동수업을 운영한 뒤 학년별과 수준별로 학업성취도를 조사한 결과 학년에 따른 학업성취도의 차이는 통계적으로 유의하지 않은 반면, 수준별로는 유의했는데 하위 분반의 성적이 지속적으로 올라간 반면 상위 분반 학습자의 성적이 하락했다고 보고하면서 수업내용과 평가내용이 일치하는지를 살펴보면서 타당한 수준별 수업의 평가방법에 대해 제안하였다. 김윤정(2015)은 서울소재 대학교에서 교양 영어를 듣는 신입생 141명을 대상으로 토익 시험으로 상중하 수준을 나누고, 13주 뒤 토익 시험으로 성취도를 평가한 결과 하위 분반이 가장 큰 상승폭을 보였다고 보고하였다. 박선호와 송현진(2013)은 의정부시 소재 초등학교 5학년 2개반을 대상으로 직전학기 성취도 평가를 기준으로 상중하 분반으로 나누어 디토크로스를 활용한 쓰기 수업을 8주간 실시하고 사후 평가한 결과 하위 분반 학생들의 쓰기 능력이 제일 많이 향상되었다고 보고하였다. 연구자들은 적은 인원수의 짧은 수업기간이 한계였지만 소규모의 수준별 수업을 지속적으로 운영한다면 수업에서 소외될 수 있는 하위권 학생들에게 긍정적인 효과를 줄 것이라고 제안하였다. 이지은과 장종득(2014)는 대구 소재 특수목적 고등학교 1학년 학생들을 대상으로 상중하 수준별과 비수준반으로 나누어 학업성취도를 조사한 결과 수준별 하반의 학생들 성적 향상도가 비수준별 반

학생들의 성적 향상도보다 높았다고 보고 하면서 수업성적이 좋지 않은 학생들로 구성된 하위반이라 할지라도 수준별 수업이 비 수준별 반에서 보다 더 높은 성적 향상을 보인 것에 주목하였다.

반면, 이화자와 정정옥(2005)는 순천시 소재 남자 중학교 1학년 학생들을 대상으로 1학기 성적으로 수준을 상하 분반으로 나누어 2학기에 수준별 이동 수업을 실시한 결과 하위 분반 학생들의 성적이 낮아졌는데 이에 대해 학생들이 ‘수업시간의 소란’을 불만족의 요인으로 지적했듯이 연구자는 하위 분반은 기존의 일반 반에서 보다 수업태도가 더 나빠지고 수업시간에 집중하지 못하는 것으로 보고하였다. 한편, 김은주와 성귀복(2016)은 수도권 소재의 초등학교 4학년을 대상으로 형태 초점 기법을 활용한 영어 문법, 읽기 및 듣기에 대한 수업을 상중하로 나누어 실시한 후 성취도 검사를 한 결과 상위 분반과 중위 분반에서 높은 향상을 보인 반면, 하위 분반에서는 변화가 없었다고 보고하였다. 정혜옥과 박부남(2014)는 서울 소재 초등학교 3-6학년 414명의 학생들을 대상으로 방과후 수준별 수업으로 학생들의 듣기, 말하기, 읽기 수업을 실시하고 변화를 조사한 결과 상위 분반에서 제일 향상도가 높았다는 결과를 보고하였다.

수준별 수업의 정의적 측면에서 학습자의 인식 및 만족도, 흥미도 등에 관한 연구 결과도 상이하게 보고되고 있지만, 본 연구의 범위를 벗어나기 때문에 여기에서는 다루지 않았다.

2. 언어의 네 가지 기능

언어 기능(language skills)이란 음성과 문자 등의 매체를 사용하여 언어 행위를 수행하는 것으로 보통 듣기(listening), 읽기(reading), 말하기(speaking), 쓰기(writing)의 네 기능으로 분류한다. 또 이 네 가지 기능을 이해 기능(receptive skills)과 표현 기능(productive skills)으로 분류하기도 하는데, 듣기와 읽기는 화자의 의도를 해석하는 정신적 과정으로 이해 기능에 해당하고, 말하기와 쓰기는 화자나 글쓴이의 표현하고자 하는 능동적 기능으로 표현 기능에 해당한다.

언어 기능은 서로 배타적인 것이 아니라 상호 보완적이어서 한 기능의 향상은 다른 기능의 향상을 촉진하며, 한 기능은 다른 기능으로 전이된다(Hughes, 1992).

네 가지 기능을 하나씩 살펴보면, 듣기는 언어 습득의 제일 기본적인 단계로, 듣기 입력에 의해서 학습자는 말을 배우게 된다. 즉, 듣기를 통하여 학습자는 다양한 수준에서 언어 체계의 상호작용에 대한 인식을 키울 수 있으며, 나아가 유창한 말하기 능력을 위한 기반을 쌓을 수 있다(Peterson, 1997). Nord(1981)는 듣기가 말하기보다 앞서야 한다고 했는데 이는 듣기가 말하기를 가능하게 하기 때문이라고 하였다.

말하기는 듣기보다 복잡한 기술로 소리와 구조, 어휘 등을 먼저 알아야 한다. 언어 학습의 궁극적 목표는 언어 형태의 습득뿐만 아니라 의사소통 기능의 숙달이다. 언어를 의사소통의 수단이라고 본다면 언어 학습의 최종 목표는 구두로 전달할 수 있는 의사소통 행위가 일어나는 상황을 알고 습득한 언어를 그 상황에서 적절히 적용하여 표현하는 데 있다(우윤주, 2006).

읽기는 문자해독(literacy)과 같은 의미로 사용될 수 있다. 일반적으로 문자 해독이란 영어 문자로 표현하고 영어 문자를 풀이하는 능력을 의미한다. 그러나 보다 엄격한 의미에서는 문자 해독 능력이란 표준 영어로 표현하는 능력과 영어를 읽고, 생각하고, 풀이하는 기술의 습득이라고 할 수 있다. 읽기란 글을 읽고 의미 이해의 단서를 추출하고, 그 단서와 자기가 이미 알고 있는 세상 지식에 근거하여 예견하며, 추측한 것을 검증하고 확인하여 다음 단계의 행동을 결정하는 활동을 말한다. 이러한 읽기 능력을 일반적으로 독해(reading comprehension) 능력이라고 한다(우윤주, 2006).

쓰기는 보통 작문(composition)과 밀접히 연결되어 있는데, 적절한 어휘나 구문을 선택하여 자기가 나타내고자 하는 의미를 표현하는 활동을 말한다. 영어 교육에서 쓰기를 가르치려는 궁극적인 목적은 영어로 자기의 생각을 글로 써서 표현하는 능력을 가르치는데 있다(조경숙, 2003).

3. 메타 분석

메타 분석은 다양한 척도를 사용한 여러 연구들을 하나의 융통성 있는 표준화된 척도로 바꾸어 효과 크기를 산출해 낸 다음 모집단의 효과 크기를 추정해내는 기법을 이용한다. 효과 크기는 효과의 양 또는 정도를 말하는 것인데, 효과 크기의 유형에는 집단간의 평균 차이(*d* family: mean differences), 변인간의 상관 계수(*r* family: correlations), 비율이나 범주형 데이터의 측정치를 포함한 승산비(categorical data: Odds ratio) 등이 있다(Borenstein, Hedges, Higgins, & Rothstein, 2009). 메타 분석으로 추정해 낸 효과 크기가 메타 분석에 쓰여진 모든 연구사례들을 아우르는 하나의 공통 평균 효과 크기가 있는지, 아니면 효과 크기의 분포로부터 평균 효과 크기를 산출해 냈는지를 알아보고(Borenstein, Hedges, & Rothstein, 2007), 추정된 효과 크기가 나타내는 의미를 파악하는 것이 메타 분석에서 하는 일이다.

메타 분석은 분석할 자료의 선정에서부터 각 연구사례¹의 효과 크기로 평균 효과 크기를 추정해가는 과정을 거친다. 우선, 분석 자료를 선정하기 위해서는 먼저 검색에 사용할 키워드를 정하고, 이 키워드들을 이용하여 데이

¹ 본 연구에서는 학술지나 학회 발표문에 실린 연구논문은 연구논문(paper)이라 칭하고, 그 연구논문에서 구체적인 효과 크기를 산출해 낼 수 있는 각각의 연구를 연구사례(study)라 칭한다.

터베이스에서 자료를 검색한다. 그리고 키워드로 수집된 모든 연구로 연구 세트의 초기 수집 목록(initial finding lists)을 정밀히 살펴보면서 포함-제외 선정 기준(inclusion-exclusion criteria)에 맞추어 자료를 포함하거나 제외할 뒤 분석할 논문의 최종 연구세트를 만든다. 이때 코딩 매뉴얼을 같이 작성하여 관련 연구 세트의 연구논문들을 코딩한다. 코딩을 할 때는 적어도 2명 이상의 코딩자가 코딩을 해야 코딩에 관한 신뢰도를 높일 수 있다.

다음, 산출된 효과 크기를 가지고 메타 분석을 할 때 몇 가지 모델을 이용해야 하는데, 이때 가중치(weights)를 달리하여 모델에 이용한다. 일반적으로, 메타 분석에서 쓰여진 관련 연구 세트는 가중화된 평균을 구하기 위해 고정 효과 가중치(fixed effects weights)를 사용하는 고정 효과 모델이나 랜덤 효과 가중치(random effects weights)를 사용하는 랜덤 효과 모델을 이용하여 분석한다. 또 혼합 효과 가중치(mixed effects weights)를 사용한 혼합 효과 모델도 있는데 이것은 조절 변인(moderator)를 이용할 때 사용한다. 모델에 사용되는 가중치는 분산의 역으로, 메타 분석에서는 두 가지의 분산이 언급된다. 연구사례 자체의 사례 수에 의해 결정되는 샘플링 에러인 연구사례 내 분산(within-study variance)과 연구사례들의 평균을 구할 때 연구사례들의 차이를 반영하는 연구사례들간 분산(between-studies variance)이다. 고정 효과 모델에 쓰여지는 고정 효과 가중치는 연구사례 내 분산의 역만 사용하는 것이고, 랜덤 효과 가중치는 연구사례 내 분산과 연구사례 간 분산을 합한 분산의 역이다(Schmidt, Oh, & Hayes, 2009). 그리고 고정 효과 모델이나 랜덤 효과 모델을 결정하기 전에 연구사례들의 모든 효과 크기가 하나의 모집단에서 나온 것인지 아닌지를 알아보는 이질성 검사가 먼저 수행되어야 한다(Schmidt et al., 2009).

III. 연구 방법

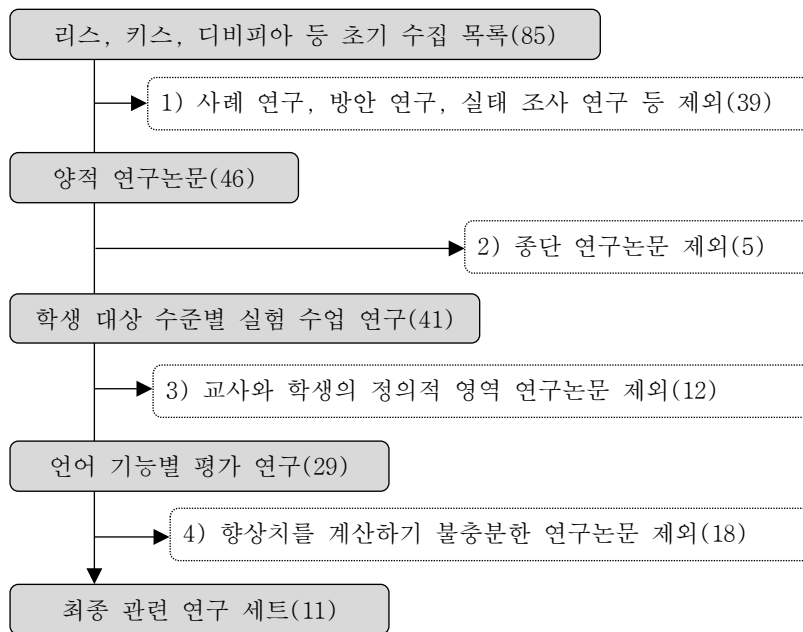
1. 메타 분석 대상 연구의 선정

수준별 영어 수업에 관한 관련 연구 세트를 위한 연구논문들을 수집하기 위하여 ‘수준별 영어 수업’, ‘수준별 초등 영어’, ‘수준별 중학교 영어’, ‘수준별 고등학교 영어’, ‘수준별 대학 영어’, ‘level differentiated English class’, ‘English ability grouping’ 등의 키워드를 사용하였다. 그리고 RISS 학술연구정보서비스, KISS 한국학술정보, DBpia, NEWnonmun, E-article 등의 데이터베이스에서 연구논문들을 검색하였다. 데이터베이스에서 논문의 제목과 초록을 살펴 수집한 초기 목록은 85편이었다.

초기 수집 목록은 포함-제외 기준에 따라 그림 1과 같은 절차로 연구논문을 선정하였다. 첫 번째, 메타 분석을 위하여 국내에서 출판된 양적 연구들

만 선정하였다. 그리고 85편의 연구 중에서 방안 연구, 연구 수업, 수업 모형 연구, 수업 설계 및 개발, 사례 연구, 평가 방안, 실태 조사와 같이 양적 연구가 아닌 연구논문 39편은 제외하였다. 두 번째, 46편의 양적 연구논문 중에서 한국교육중단연구 자료를 사용하여 다층 분석을 한 연구 5편을 제외하고 수준별 수업에 관해 실험 연구를 한 논문 41편을 선정하였다. 세 번째, 수준별 수업에 관한 학생 인식, 교사 인식 등과 같은 정의적 영역의 변화만을 수록한 연구논문 12편을 제외하였고, 마지막으로 메타 분석을 위한 충분한 데이터(예를 들면, 수준별 집단의 사례 수, 평균과 표준 편차 또는 사전 사후 대응 표본 t 검정의 결과로 t 값이나 p 값)가 수록되지 않은 논문 18편을 제외하고 최종 11편의 논문을 메타 분석 대상으로 선정하였다. 11편의 논문에서 사전사후 향상도인 평균 차이를 나타내는 d 효과 크기와 분산을 산출해 낼 수 있는 총 1,498명을 대상으로 한 55개의 연구사례를 선정하여 메타 분석하였다.

그림 1
관련 연구 세트 선정 절차



2. 코딩 절차와 데이터 정리

메타 분석을 위한 코딩과 데이터 정리를 위하여, 본 연구에서는 두 명이 코딩을 담당한 코딩자의 역할을 하였으며 개별적으로 연구논문의 내용적 측

면과 효과 크기 특성 측면에 관하여 코딩하였다. 연구논문에 관한 기본적인 사항인 저자, 제목, 출판 연도, 출판 상태, 저널명과 더불어 논문의 내용 측면에서는 분반 형태, 학교 급, 학년, 수준별 수업의 내용, 수준을 구분한 방법, 그리고 어떤 언어 기능을 평가하였는지 코딩하였다. 효과 크기 특성 측면에서는 수준에 따라서 사례 수와 평균, 표준 편차, 사전사후 평균 차이, 편차 차이, t 값, p 값과 더불어 본 연구에서 사용한 데이터 추출 방법과 데이터가 있는 페이지나 표의 번호를 코딩하였다. 그리고 두 코딩자의 코딩 일치도, 즉 코딩 신뢰도는 코딩 항목(변인)에 따라 11편의 연구논문을 퍼센트로 알아보았다. 두 명의 코딩자는 언어 기능 변인에 60%, 수업 내용 변인에 73%, 그리고 학년 구분과 수준 구분은 93%의 일치율을 보였으나 나머지 변인에 관해서는 100%의 일치율을 보였다. 그래서 완전히 합의가 될 때까지 의논한 뒤 11편의 연구논문을 본 연구의 관련 연구 세트에 선정하였다. 따라서 총 11편의 연구논문에서 55개의 연구사례가 도출되었으며 본 메타 분석에 쓰일 d 효과 크기와 분산을 계산하였다. 본 연구에 조절 변인으로 사용된 변인에 대한 자세한 내용은 표 1과 같다.

표 1
조절 변인의 코딩 값, 연구사례 수

변인	수준 구분	분반형태	학교 급	학년	언어 기능
코딩 값	사전평가(25)*	상(19)	초등학교(31)	초등 4학년(8)	듣기(18)
	성취결과(26)	중(16)	중학교(7)	초등 5학년(14)	말하기(9)
	학생선택(4)	하(20)	고등학교(5)	초 3-6학년(9)	읽기(8)
			대학교(12)	중 1학년(5)	쓰기(2)
				중 2학년(2)	읽기-듣기(2)
				고 1학년(5)	어휘(2)
				대 1학년(6)	문법(5)
			대 1-4학년(6)	학업성취도(9)	

* 괄호 안의 수는 연구사례 수.

본 연구에 선정된 11편의 연구논문은 2003년부터 2017년까지 출판되었으며 모두 학술지에 게재되었다. 수준별 수업은 원어민 활용 수업, 이동 수업(나기연 & 정미란, 2003; 정혜옥 & 박부남, 2014) 등 다양한 내용으로 이루어져 있었다. 그리고 수준을 구분하는 방법은 사전평가를 하거나 전 학기 학업 성취 결과나 기본 과정의 이해 정도에 따라서 구분하였으며, 학습자의 선택이나 희망(이지은 & 장종득, 2014; 이화자 & 정정옥, 2005)으로 구분하기도 하였다. 그리고 결과로 나온 언어 기능 측면은 듣기, 말하기, 읽기, 쓰기, 읽기-듣기, 어휘, 문법 영역과 학업 성취도 평가로 사전사후 향상도를 보여주었다. 또한 주로 대응표본 t 검정과 아노바 등으로 평균 차이를 도출해 냈으며 본 연구에서는 기술 통계치와 t 값을 이용하여 d 효과 크기를 계산하였다.

3. 효과 크기와 분산의 계산

본 연구의 효과 크기는 사전사후 평가의 향상도, 즉 사전사후 평가의 평균 차이를 결과측정치(outcome measures)로 하였다. 따라서 본 연구는 d 효과 크기가 결과측정치다. 메타 분석을 하기 위해 효과 크기와 그 효과 크기의 분산을 구하였다. 11편의 연구논문에서 나온 55개 연구사례의 사례 수와 t 값으로 방정식(1)과 같이 각 연구사례의 편향 표준화 평균 차이(biased standardized mean differences, g)를 구하였다. 그 다음 방정식(2)의 교정 함수 J 를 방정식(3)과 같이 곱하여 각 연구사례의 불편 표준화 평균 차이(unbiased standardized mean differences)인 d 효과 크기를 계산하였다(Lin, 2018; Morris & DeShon, 2002; Novianti, Roes, & der van Tweel, 2014). 사전사후 평가의 원점수(raw score)로 효과 크기를 계산한다면 편향되지 않지만, 사전사후 평가의 향상치로 효과 크기를 계산하면 편향되기 때문에 (Morris & DeShon, 2002), 이것을 교정하는 교정 함수 J 를 이용하여 각 연구 사례들이 가지고 있는 편향성을 교정하였다.

$$g = \frac{t}{\sqrt{n}} \quad (1)$$

$$J = 1 - \frac{3}{(4df-1)}, \quad (2)$$

$$d = g \times J \quad (3)$$

여기에서 t 값은 사전사후 대응 표본 t 검정의 t 값이고, n 은 사전사후 대응 표본의 사례 수이며, 사전사후 대응 표본 검사에서 자유도(df)는 $n-1$ 이다. 그리고 d 효과 크기의 분산은 방정식(4)와 같이 계산하였다

$$Var(d) = \frac{1}{n} \times \frac{(1+g^2)}{2} \quad (4)$$

4. 데이터 분석 방법

본 연구에서는 메타 분석을 위하여 R 소프트웨어 버전 3.6.1(R Core Team, 2019)의 *metaphor* 패키지에 있는 *rma*와 *forest*, *trimfill*, *funnel*기능을 이용하여 분석하였다(Viechtbauer, 2010).

우선, 가장 적절한 모델을 결정하기 위하여 연구의 이질성 검사(tests of heterogeneity 또는 Q tests)를 하였다. 즉, 메타 분석에 사용할 모든 연구들이 하나의 모집단에서 비롯된 것인지를 알아보기 위해 이질성 검사를 하였다. 연구사례들이 이질적인지를 판단하기 위해 Q 값과 Birge 비율(Birge's ratio), I^2 값으로 살펴보았다. Q 값과, Q 값을 df 로 나눈 Birge비율은 연구사례의 사례 수에 민감하다. 반면에 이질성의 정도를 나타내는 I^2 값은 앞의 두

지수와는 달리 퍼센트로 나타내는 지수로 전체 분산에 대한 연구사례간의 분산이 차지하는 비율이다. 따라서 I^2 값으로 메타 분석에 쓰여진 연구사례들이 얼마나 이질적인지를 판단할 수 있으며 서로 비교도 가능하다(Higgins, Thompson, Deeks, & Altman, 2003). Q 값과 Birge 비율은 그 값이 크면 클수록 연구사례들이 이질적이라는 것이지만, 이질적이라고 판단하는 기준치(cut off)는 없다. 하지만 Q 검증을 통해 p 값도 같이 산출됨으로 이질적인지 아닌지를 판단한다. Higgins 등이(2003)가 정한 기준에 의하면, I^2 값이 25% 이하면 이질성이 낮고 75% 이상이면 아주 이질적이다.

이질성 검사 후 평균 효과 크기를 추정하였다. 그리고 조절 변인인 상중하 분반, 학교 급, 학년, 언어 기능, 수준 구분 방법에 따라 평균 효과 크기를 추정하기 위하여 혼합 효과 모델(mixed-effects model)을 사용하였고, 마지막으로 출판 편향을 살펴보았다.

IV. 연구 결과

1. 이질성 검사 결과(Q tests)

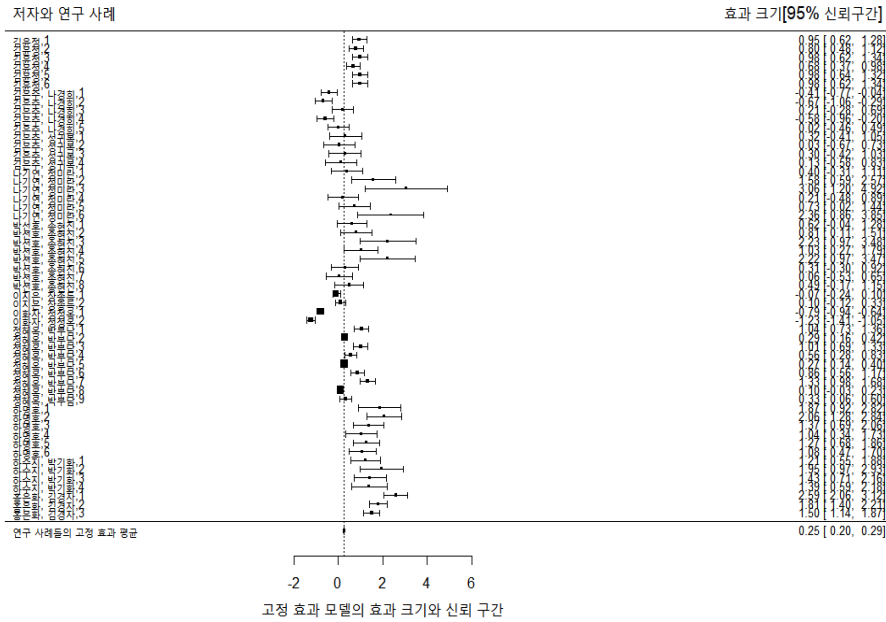
표 2는 사전사후 향상치를 나타내는 효과 크기의 모집단 분산이 0이 아님을 보여준다. 즉, 이 메타 분석에 사용된 연구사례들은 하나의 모집단에서 나온 것이 아니라 이질적 모집단을 가진 여러 연구사례들이라는 것이다. 통계적으로 유의한 이질성 검사($Q(54) = 1042.83, p < .001$) 결과와 더불어 Birge 비율(19.31)과 I^2 값(95.3%)은 본 메타 분석의 연구사례들이 큰 편차를 가졌다는 것을 일관성 있게 보여준다.

표 2
이질성 검사 결과

Q	df	p	Birge's ratio	I^2
1042.83	54	< .001	19.31	95.3%

또한 본 메타 분석에 사용된 55개 연구사례의 d 효과 크기와 신뢰 구간을 저자 순으로 나열한 그림 2의 숲 도표(Forest plot)는 이질성 검사의 결과를 한 눈에 시각적으로 보여주는 도표이다. 숲 도표는 연구논문의 저자와 연구사례에 따라 효과 크기와 신뢰 구간을 점과 선으로, 그리고 효과 크기와 95% 신뢰 구간의 수치를 한 눈에 보여준다. 따라서 그림 2는 표 2의 이질성 검사와 마찬가지로 본 연구사례들이 이질적이라는 것을 시각적으로 뒷받침해 준다. 따라서 본 메타 분석은 랜덤 효과 모델을 이용한 랜덤 효과 평균(random-effects mean)으로 추정하였다.

그림 2
 숲 도표



2. 랜덤 효과 모델 결과

표 3의 고정과 랜덤 효과 모델 결과에는 고정 효과 모델과 랜덤 효과 모델의 평균 효과 크기 추정치(estimate)와 표준 오차(standard error), z값, 95% 신뢰 구간의 상한선과 하한선, p값을 포함시켰다. 추정된 모집단의 사전사후 향상도 평균 효과 크기는 0.74이고 추정된 평균 효과 크기의 평균 오차는 .111이며, 통계적으로 유의한 결과를 보였다($p < .001$). Cohen(1988)의 d 효과 크기 해석 기준에 따르면 ‘0.20 이하’의 효과 크기는 ‘작은 효과 크기’, ‘0.50’은 ‘중간 효과 크기’, ‘0.80이상’은 ‘큰 효과 크기’로 해석한다고 하였다. 그러므로 본 연구의 랜덤 효과 평균 0.74는 큰 효과 크기라고 할 수 있으며 이것은 수준별 영어 수업의 사전사후 향상도가 크다는 것을 보여주는 것이다. 따라서 국내에서 이루어진 수준별 영어 수업은 학습자들의 언어 능력 측면에 효과가 있다고 할 수 있다. 랜덤 효과 분석의 결과에서 보듯이 연구사례간에 차이가 있었기 때문에 다음은 조절 변인 분석(moderator analysis)을 통해 어떤 조절 변인이 사전사후 향상도에 영향을 주었는지를 알아보았다.

표 3
고정과 랜덤 효과 모델 결과

모델	평균	표준 오차	z	95% 신뢰구간		p
				하한값	상한값	
고정 효과	.25	.022	11.07	.20(-.29)	.29(.74)	< .001
랜덤 효과	.74	.111	6.65	.52(.32)	.96(.16)	< .001

3. 조절 변인 분석

조절 변인 분석은 예측 변수(predictor), 즉 조절 변인인 수준별 분반 형태, 학교 급, 학년, 언어 기능, 수준 구분 방법을 통해 사전사후 향상치를 예측할 수 있는지 분석하였다. 여기에서 모든 조절 변인이 범주형 변인(categorical variable)이기 때문에 ANOVA(Analysis of Variance)와 유사한 범주형 모델(categorical models)을 이용한 혼합 효과 분석을 하였다. 혼합 효과 모델의 조절 변인에 대한 영가설은 ‘조절 변인의 모든 범주(예를 들면, 수준 조절 변인의 상, 중, 하 범주)의 모집단 평균이 같다’이다. 이것은 $Q_{between}$ 을 검증하여 유의하면 ANOVA의 전반적 검정 결과 해석처럼 적어도 어느 한 범주는 또 다른 한 범주와 통계적으로 유의한 차이가 있다고 보는 것이다. 따라서 전반적 검정이 유의하면 조절 변인의 범주 별로 유의한 차이가 있는지 대조 분석(contrast analysis)도 시행하였다. 반면, Q_{within} 또는 Q_{error} 는 의미가 없기 때문에 기록하지 않았다(Wilson, 2011).

1) 분반 형태별 조절 변인

표 4
분반 형태 조절 변인의 혼합 효과 모델 결과

분반 형태	사례 수	평균효과 크기 추정치	표준 오차	p
상	19	.64	.187	< .001
중	16	.59	.203	< .001
하	20	.97	.190	< .001
분반 형태 조절 변인 $Q_{between}(2) = 2.31, p < .315$				

종속 변인인 결과측정치는 d 효과 크기이고, 조절 변인은 범주형 변인으로 범주는 상, 중, 하 3개 수준이다. 먼저, 수준별로 d 효과 크기의 평균이 범주 별로 서로 같은지를 알아본 결과, 표 4와 같이 수준 조절 변인은 통계적으로 유의하지 않았다($Q_{between}(2) = 2.31, p = .315$). 즉, 상중하 수준은 사전사후 향상치인 d 효과 크기에 대한 중요한 예측 변인이 아니었다. 하지만, 상중하 수준자체의 평균 효과 크기 추정치는 각 수준별로 통계적으로 유의하였다($\overline{d}_{상}^2 = 0.64, \overline{d}_{중}^2 = 0.59, \overline{d}_{하}^2 = 0.97$, 모두 $p < .001$). 즉, 수준간 평균

효과 크기의 차이는 통계적으로는 유의하지는 않지만, 수준별 효과 크기는 통계적으로 유의하며 대체적으로 보통 이상의 효과가 있었다. 또한, 수준별 수업은 하위 집단에게 제일 효과가 크고, 다음으로 상위 집단, 그리고 중위 집단이었다.

2) 학교 급 별 조절 변인

학교 급 별 조절 변인도 초등학교, 중학교, 고등학교, 대학교로 구성된 범주형 변인이라서 ANOVA와 유사한 혼합 효과로 분석하였다. 표 5와 같이 학교 변인은 통계적으로 유의하기 때문에($Q_{between}(3) = 40.61, p < .001$), 학교 조절 변인은 학교간 d 효과 크기에 대한 중요한 예측 변인이라고 할 수 있다. 또한, 초등학교와 고등학교, 대학교는 99.99%수준에서 사전사후 향상치 효과 크기 자체가 통계적으로 유의하였다($\overline{d}_{\text{초}} = 0.77, \overline{d}_{\text{고}} = 1.12, \overline{d}_{\text{대}} = 1.12$, 모두 $p < .001$). 특히 고등학교와 대학교 학생들의 사전사후 향상도가 아주 큰 효과 크기를 보여주었다. 즉, 수준별 수업은 고등학교와 대학교에서 효과가 크다고 할 수 있다.

표 5
학교 급 별 조절 변인의 혼합 효과 모델 결과

학교	사례 수	효과 크기 추정치	표준 오차	p
초등학교	31	.77	.115	< .001
중학교	7	-.52	.216	.016
고등학교	5	1.12	.256	< .001
대학교	122	1.12	.175	< .001
학교 조절 변인 $Q_{between}(3) = 40.61, p < .001$				

다음은 어느 학교와 어느 학교의 사전사후 향상도가 통계적으로 유의한 차이가 있는지 대조 분석을 하였다. 대조 분석은 z 통계치를 계산하여 임계치(critical value)인 1.96보다 크면 검사한 대조 쌍이 통계적으로 유의하게 차이가 있다고 해석하였다. 그 결과, 중학교와 대학교($z = 5.89$), 초등학교와 중학교($z = 5.28$), 중학교와 고등학교($z = 4.89$)가 통계적으로 유의한 차이를 보였다. 정리하면, 수준별 영어 수업은 고등학교와 대학교 학생들의 사전사후 향상도를 가장 많이 높이는 수업이라고 할 수 있고, 중학교 학생들에 비해 초등학교, 고등학교, 대학교 학생들의 영어 향상도를 높여주는 수업이라고 할 수 있다.

² $\overline{d}_{\text{상}}$ 는 상 그룹의 d 효과 크기의 평균 추정치를 나타낸다.

3) 학년 조절 변인

학년 조절 변인은 초등학교 4학년, 5학년, 3학년부터 6학년, 중학교 1학년, 2학년, 고등학교 1학년, 대학교 1학년, 그리고 대학교 1학년부터 4학년의 8개의 범주로 이루어졌다. 유사 ANOVA 혼합 효과 분석 결과는 표 6에서 보는 것처럼 학년 변인도 학년간 d 효과 크기에 대한 중요한 예측 변수라는 것을 보여준다($Q_{between}(7) = 47.19, p < .001$). 그리고 초등학교 4학년과 5학년, 초등학교 3-6학년, 고등학교 1학년, 대학교 1학년, 대학교 1학년-4학년은 사전사후 향상도 효과 크기 자체가 통계적으로 유의하였다($\overline{d}_{z4} = 0.81, \overline{d}_{z5} = 0.89, \overline{d}_{z3-6} = 0.63, \overline{d}_{z1} = 1.11, \overline{d}_{z1-4} = 0.89, \overline{d}_{z1-4} = 1.42, p < .001$). 또한, 대조 분석 결과, 중학교1학년과 2학년은 초등학교4학년(각각 $z = 3.19, z = 4.06$), 초등학교5학년(각각 $z = 3.76, z = 4.47$), 초등학교3-6학년(각각 $z = 2.94, z = 3.86$), 고등학교1학년(각각 $z = 3.92, z = 4.63$), 대학교1학년(각각 $z = 3.46, z = 4.26$), 대학교1-4학년(각각 $z = 4.65, z = 6.64$) 모두와 통계적으로 유의한 차이가 있었으며, 초등학교 3-6학년은 대학교 1-4학년($z = 2.44$)과 통계적으로 유의한 차이를 보였다. 학년 조절 변인별로 살펴본 결과, 대학교 1학년-4학년에서 수준별 수업의 효과가 제일 컸고, 다음으로 고등학교 1학년, 그리고 대학교 1학년, 초등학교 5학년과 초등학교 4학년 순으로 효과가 크게 나타났다. 전반적으로 수준별 수업은 중학교 학생을 제외한 초, 고, 대학교 학생들에게 효과가 큰 것으로 나타났다.

표 6
학년 조절 변인의 혼합 효과 모델 결과

학년	학년 연구사례 수	효과 크기 추정치	표준 오차	p
초 4학년	8	0.81	.233	< .001
초 5학년	14	0.89	.186	< .001
초 3-6학년	9	0.63	.258	< .001
중 1학년	15	-0.30	.383	.246
중 2학년	2	-1.01	.253	.008
고 1학년	5	1.11	.229	< .001
대 1학년	6	0.89	.230	< .001
대 1-4학년	6	1.42	.265	< .001

학년 조절 변인 $Q_{between}(3) = 40.61, p < .001$

4) 언어 기능 조절 변인

언어 기능 조절 변인은 듣기, 말하기, 읽기, 쓰기, 듣기-읽기, 어휘, 문법, 성취도로 구성된 8개의 범주로 이루어졌다. 표 7의 유사 ANOVA 혼합 효과 분석 결과에 의하면 언어 기능 변인은 사전사후 향상도의 효과 크기가 통계적

으로 유의한 차이를 보이지 않았다($Q_{between}(7) = 9.25, p = .235$), 단지 듣기, 말하기, 읽기 자체의 사전사후 향상도 효과 크기가 통계적으로 유의하였고($p < .01$), 효과 크기도 큰 것으로 나타났다. 그리고 듣기와 읽기, 말하기 순으로 효과가 컸다($\bar{d}_{\text{듣기}} = 0.88, \bar{d}_{\text{읽기}} = 0.87, \bar{d}_{\text{말하기}} = 0.85$). 따라서 수준별 수업은 듣기, 말하기, 읽기에 효과가 있다고 할 수 있다.

표 7

언어 기능 조절 변인의 혼합 효과 모델 결과

언어 기능	수준 연구사례 수	효과 크기 추정치	표준 오차	p
듣기	18	.88	.195	< .001
말하기	9	.85	.273	.002
읽기	8	.87	.278	.002
쓰기	2	1.53	.642	.017
듣기-읽기	2	.17	.591	.768
어휘	2	1.55	.609	.011
문법	5	.26	.368	.487
성취도	9	.34	.257	.188

학년 조절 변인 $Q_{between}(3) = 9.25, p = .235$

5) 수준 구분 조절 변인

수준 구분은 사전평가, 직전 학기 성취 결과, 그리고 학생의 희망이나 선택으로 이루어졌다. 표 8과 같이 수준 구분 조절 변인도 통계적으로 좋은 예측 변인이었다($Q_{between}(2) = 17.54, p = .0002$). 분석 결과 사전평가와 직전 학기 성취 결과($\bar{d}_{\text{사전}} = 0.98, \bar{d}_{\text{성취}} = 0.68, p < .001$)로 수준을 구분했을 때 사전사후 향상도가 통계적으로 유의하게 ($p < .001$) 큰 효과를 보여주었으나, 학생의 선택이나 희망으로 수준을 구분하여 수준별 수업을 했을 때는 통계적으로 유의한 향상을 보이지 않았다. 또한 대조 분석 결과에서도 사전평가나 직전 학기 성취에 의한 수준 구분은 학생의 선택이나 희망에 의한 수준 구분과 각각 통계적으로 유의한 차이를 보였다($z = 4.17, 3.29$). 따라서 사전평가에 의한 수준 구분이 제일 효과가 크고, 다음이 직전 학기 성취도에 의한 구분이었다. 학생의 선택에 의해 수준을 구분하였을 때는 사전사후 향상도가 없었다.

표 8

수준 구분 조절 변인 혼합 효과 모델 결과

수준 구분	연구사례 수	추정치	표준 오차	p
사전평가	25	.98	.139	< .001
이전 성취 결과	26	.68	.146	< .001
학생 선택	4	-.50	.327	.125

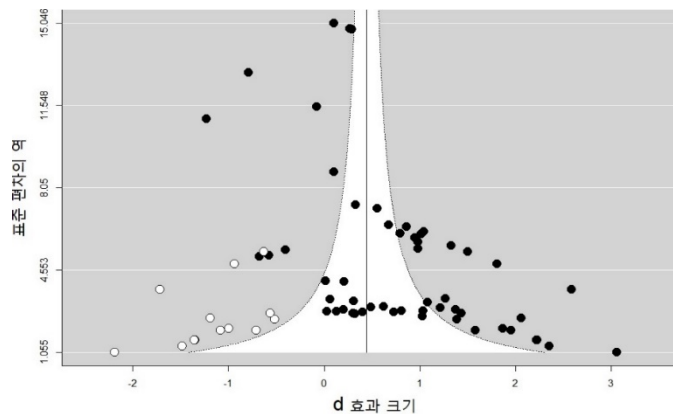
학년 조절 변인 $Q_{between}(3) = 9.25, p = .235$

4. 출판 편향(publication bias) 결과

출판 편향은 서랍 속에 감춰진 효과 또는 문제라고도 하는데, 연구자가 세운 영가설을 기각하지 못해, 즉 효과 있는 결과를 도출해 내지 못해 학술지나 학회에 게재나 발표되지 못하고 연구자의 서랍 속에 숨겨둔 연구논문이 있다는 생각에서 나온 용어이다(Rosenthal, 1979). 이것은 연구자의 가설을 지지하거나 효과가 있는 연구들만 게재 또는 발표되어 학계에 나오기 때문에 출판에 있어서 편향이 있다는 것이다. 그래서 메타 분석에 선정된 관련 연구 세트의 연구논문들이 출판 편향이 있는지 없는지 검증해야 한다. 본 연구에서는 출판 편향을 시각적으로 알아볼 수 있는 깔때기 도표(funnel plot)와 통계적 검사인 다듬고 채우기(trim and fill) 기법으로 알아보았다.

그림 3 는 깔때기를 거꾸로 엮어 놓은 모양으로 가로축은 d 효과 크기를, 세로축은 표준 편차의 역을 나타낸다. 그리고 깔때기 도표의 가운데 세로선은 랜덤 효과 평균인 0.74 를 나타낸다. 깔때기는 평균을 중심으로 좌우 대칭을 이루며 검은색 점은 본 메타 분석에 사용된 연구사례들의 효과 크기와 편차에 맞춰 찍힌 것이고, 흰색 점은 본 메타 분석에 결여된 연구사례를 표시한 것이다. 따라서 본 메타 분석에서는 중앙에 있는 평균 효과 크기의 왼쪽 아래 부분에 13 개의 연구사례가 결여되었다고 나왔으며 만약에 13 개의 연구사례를 포함한 68 개의 연구사례로 메타 분석을 한다면 랜덤 효과 평균이 0.44 가 될 것이라고 예측하였다. 그림 3 으로 효과 크기가 작거나 또는 표본의 크기가 큰 논문들이 숨어 있을 것이라는 것을 예상할 수 있다.

그림3
깔때기 도표



따라서 본 메타 분석은 출판 편향이 있다고 할 수 있는데, 이것은 본 연구에 쓰여진 연구 논문들이 모두 학술지에 게재된 연구논문이기 때문이라고 생

각한다. 보통 학술지에 게재되는 연구논문들은 통계적으로 유의하고 효과 크기가 큰 논문들이기 때문에 통계적으로 유의하지 않거나 효과 크기가 작은 논문들은 학술지에 등재되지 못했을 것이다. 동시에 이것이 본 논문의 한계점이라고도 할 수 있다.

V. 결론 및 제언

학습자의 개인차에 상관없이 모두에게 적합한 교육을 제공한다는 수준별 수업의 교육 철학과 달리 수준별 교육의 효과에 대한 논쟁(Abraham, 1989; Ansalone, 2001; Gamoran & Mare, 1989; Kulick & Kulick, 1982; Oakes, 1990; Wilson & Ribovich, 1973)과 상이한 연구 결과(강보람 & 조윤경, 2012; 김윤정, 2015; 김은주 & 성귀복, 2016; 박선호 & 송현진, 2013; 이화자 & 정정옥, 2005; 정혜옥 & 박부남, 2014)가 지속적으로 보고되고 있다. 이에 본 연구는 2차 연구로서의 메타 분석 기법을 적용하여 수준별 수업의 효과가 어느 정도인지를 체계적이고 객관적인 입장에서 조망해 보았다.

본 연구에서 수준별 영어 수업에 관한 언어의 네 가지 기능에 관한 연구들 중 메타 분석에 필요한 데이터가 충분하다고 판별된 최종 11편의 연구논문을 통해 55개의 개별 연구사례들을 추출하였고, 이는 2003년부터 2017년 사이에 발표된 연구 논문들로 1,498명을 대상으로 하였다. 이 대상들을 가지고 수준별 영어 수업의 전반적인 학습 효과와 세부적으로는 분반 형태, 학교 급, 학년, 언어 기능, 수준 구분 방법에 따라 학습 효과를 분석하는 연구를 통해 다음과 같은 결론을 도출하였다.

첫째, 수준별 수업의 전체 효과 크기는 0.74로 ‘큰 효과’ 크기로 나타나 수준별 영어 수업이 EFL 학습자의 영어 능력을 향상시키는데 효과적인 교수방법임을 말해준다. 이는 Kulick와 Kulick(1982)가 ESL 학습자를 대상으로 한 연구에서 보고된 ‘중간 효과’ 보다 큰 것으로 나왔기 때문에 수준별 영어 수업은 ESL 학습자에 비해 EFL로서의 국내 영어 학습자에게 더 효과가 있는 것인지는 추후 연구를 통해 더 살펴볼 문제이다.

둘째, 분반 형태에 따른 수준별 수업의 효과 크기는 하위 집단($\bar{d} = 0.97$), 상위 집단($\bar{d} = 0.64$), 그리고 중위권 집단($\bar{d} = 0.59$)의 순으로 나타나긴 했지만, 통계적으로 유의한 차이는 아니었다. 그러나 EFL 영어 학습자를 대상으로 한 수준별 수업은 하위권학생에게 도움이 된다고 할 수 있으며, 하위권의 영어 향상도에 도움이 되었다는 선행 연구(강보람 & 조윤경, 2012; 김영미, 2010; 김윤정, 2015; 박선호 & 송현진, 2013; 이지은 & 장종득, 2014; 정양수, 2014; 하수지 & 박기화, 2017)와도 맥을 같이한다.

셋째, 학교 급 별로 살펴본 결과, 수준별 수업의 효과 크기는 고등학교($\bar{d} = 1.12$), 대학교($\bar{d} = 1.12$), 초등학교($\bar{d} = 0.77$)로 중학교를 제외하고 초등

학교, 고등학교, 대학교에서 시행할 때 효과가 크게 나타났다. 또한, 학년별로 살펴본 결과도 같은 맥락으로 해석될 수 있는데, 대학교 1학년-4학년($\bar{d} = 1.42$), 대학교 신입생($\bar{d} = 0.89$), 고등학교 1학년($\bar{d} = 1.12$), 초등학교 5학년($\bar{d} = 0.89$), 4학년($\bar{d} = 0.81$), 그리고 초등학교 3학년-6학년($\bar{d} = 0.63$)을 대상으로 한 수준별 수업이 모두 효과가 좋은 것으로 나타났다.

넷째, 언어 기능별로 살펴보았을 때, 듣기($\bar{d} = .88$), 읽기($\bar{d} = .87$), 그리고 말하기($\bar{d} = .85$)의 순서로 효과가 큰 것으로 나타났으나 통계적으로 유의하지는 않았다.

다섯째, 분반을 구분한 방법에 따라 살펴보았을 때, 수준 구분을 사전평가($\bar{d} = 0.98$), 그리고 직전 학기 성취 결과($\bar{d} = 0.68$)로 했을 때 효과가 있었고, 특히, 사전평가 방식이 효과가 매우 컸다. 학생의 선택이나 희망으로 수준을 구분하였을 때는 수준별 수업의 효과가 없었다.

마지막으로 본 연구에 사용된 논문들은 출판 편향을 보이고 있는데 이는 본 연구에 사용된 논문들이 모두 학술지에 게재된 연구논문이기 때문일 것이고 이것이 또한 본 연구의 제한점이기도 하다. 또 다른 제한점으로 본 연구 결과 중학교 1학년과 2학년을 대상으로 한 수준별 영어 수업의 효과가 가장 작았으며, 통계적으로 유의하지 않았다. 그 이유는 중학교 2학년을 대상으로 한 연구에서는 수준을 학생의 선택이나 희망에 따라 구분하였고, 사전사후 향상도 결과를 학업 성취도 평가로 하였기 때문이라고 분석된다. 반면, 대학생을 대상으로 한 연구는 모두 사전평가의 결과에 따라 수준을 구분하였다. 따라서 본 연구에서 학년과 수준 구분 방법이 서로 구분되지 않고 교란된 결과일 수도 있다.

본 연구를 바탕으로 효율적인 수준별 영어 교육 방법을 위한 다음과 같은 제언을 하고자 한다. 첫째, 수준별 수업의 효과를 보다 종합적으로 살펴보기 위해서 1차 연구가 더 많이 이루어져야 한다. 예로 중학생을 대상으로 한 연구사례는 충분하지가 못하였다. 둘째, 출판 편향을 해소하기 위하여 서랍속의 논문이라고 불리는 다양한 형태의 논문들을 포함시켜야 한다. 셋째, 1차 연구자들이 연구의 설계 및 수행 절차, 그리고 연구 결과를 보고하는데 보다 세심한 연구 설계, 통계 분석과 성실한 자세가 요구된다. 예를 들면, 수준 결정을 학생 임의로 하게 하거나, 사전사후 향상도 평가를 하지 않거나 하면 객관적이고 정확한 연구 결과를 알 수 없다. 마지막으로, 후속 연구로 학습자의 흥미, 만족도, 자신감 등 학습자의 정의적 영역 및 인식에 대한 효과를 알아보면 수준별 수업 효과에 대한 보다 다각적인 분석이 가능할 것으로 본다. 이러한 연구를 통해 수준별 수업을 계획하고 개선하는데 더 많은 객관적인 정보와 기초자료를 제공할 수 있을 것이다.

참고문헌

*표시는 메타 분석 대상 논문임.

- 강보람, 조윤경. (2012). 중학교 영어 과 수준별 수업이 학업성취도에 미치는 영향. *현대영어교육*, 13(2), 101-120.
- 권영은, 박은수. (2011). 수준별 영어수업에 대한 중학생의 만족도 탐색, *Studies in English Education*, 16(2), 1-23.
- 김영미. (2010). 대학 영어의 수준별 교육에 관한 실험 연구. *영어교육연구*, 15(1), 49-76.
- 김영숙, 이지연. (2009). 대학 교양영어 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), 87-103.
- *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.
- *나기연, 정미란. (2003). 초등영어교육에서 영어 원어민을 활용한 수준별 수업의 효과. *영어교육연구*, 8(1), 131-152.
- 민지아, 조윤경. (2011). 고등학교 수준별 영어 수업에 대한 인식 조사: 영어 학습 부진아 및 영어 교사를 대상으로. *중등영어교육*, 7(3), 75-98.
- 박상욱. (2004). 효과적인 중학교 수준별 영어 수업 방안에 대한 연구. *한국교원교육연구*, 21(3), 71-89.
- *박선호, 송현진. (2013). 디кто글로스 활용 수준별지도가 초등학생 영어능력 신장 및 정의적 영역에 미치는 효과. *응용언어학*, 29(4), 249-288.
- 박지영, 오준일, 이인규. (2012). 대학의 수준별 토익 수업에 대한 학습자의 만족도에 관한 연구. *영어교육연구*, 24(3), 257-276.
- 박찬규. (2013). 대학 교양영어 수준별 수업의 효율성 제고 방안 연구: 학습자와 교수자의 인식도를 중심으로. *언어학 연구*, 10, 105-131.
- 심지현, 임현우. (2017). 학습자 변인 및 분반 형태에 따른 영어 과 수준별 수업의 효과. *중등영어교육*, 65(4), 883-912.
- 오성삼. (2002). *선행연구 결과의 통합과 재분석을 위한 메타 분석의 이론과 실제*. 서울: 건국대학교 출판부.
- 우윤주. (2006). *언어기능 향상을 위한 동화 재즈 챗트 적용 초등영어 프로그램 개발*. 미출간 석사학위 논문. 이화여자대학교, 서울.

- 이경화, 김경혜, 한문섭. (2016). 영어 수준별 이동수업이 초등학교 학생들의 정의적 영역에 미치는 영향 및 개선 방안. *학습자중심교과교육*, 16(2), 667-692.
- 이제영, 문은주, 박옥희. (2013). 영어수업에서의 영화 활용의 효과: 국내 연구에 대한 메타 분석. *영상영어교육*, 14(3), 87-106.
- 이제영, 박옥희. (2016). 온라인을 이용한 대학 영어 교육 효과에 대한 메타 분석. *외국학연구*, 37, 69-95.
- *이지은, 장종득. (2014). 수준별 영어 수업이 학업 성취에 미치는 영향. *언어과학연구*, 71, 197-216.
- *이화자, 정정옥. (2005). 학습태도와 학습성취에 미치는 수준별 이동수업의 효과. *현대영어교육*, 6(2), 150-181.
- 정양수. (2014). 교양영어 수준별 수업의 운영 효과 및 개선 방안에 관한 연구. *언어학연구*, 19(3), 127-148.
- *정혜옥, 박부남. (2014). 방과후 초등학교 수준별 영어학습 참여의 효과. *열린교육연구*, 22(1), 1-20.
- 조경숙. (2003). *초등영어 읽기 쓰기 교육*. 서울: 한국문화사.
- 하명애. (2018). 원어민교수의 교양영어 수준별 수업에 대한 학습자와 교수자의 인식 연구. *현대영어영문학*, 62(1), 223-250.
- *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.
- 황여정. (2008). 수준별 이동수업이 학업 성취에 미치는 영향. *한국과학연구*, 14(1), 71-101.
- Abraham, J. (1989). Testing Hargreaves' s and Lacey' s differentiation-polarisation theory in a settled comprehensive. *British Journal of Sociology*, 40(1), 46-81.
- Ansalone, G. (2001). Schooling, tracking and inequality. *Journal of Children and Poverty*, 7(1), 33-49.
- Ball, S. (1981). *Beachside comprehensive: A case study of secondary schooling*. Cambridge, MA: Cambridge University Press.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. John Wiley & Sons. doi: 10.1002/9780470743386.

- Borenstein, M., Hedges, L. V., & Rothstein, H. R. (2007). *Meta-analysis fixed effect vs. random effects*. Retrieved October 10, 2017, from Meta-Analysis.com Website: 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.). New York: Longman.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- 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.
- 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.
- Hughes, A. (1992). *Testing for language teachers*. Cambridge, UK: Cambridge University Press.
- Kromrey, J., & Hogarty, K. (1963). Estimates of variance components in random effects meta-analysis: Sensitivity to violations of normality and variance homogeneity. In American Statistical Association Joint Statistical Meeting, 1963-1968.
- Kulick, J. A. (1992). An analysis of the research on ability grouping: Historical and contemporary perspectives. *Ability Grouping*, 1-45.
- Kulick, C. L., & Kulick, J. A. (1982). Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings. *American Educational Research Journal*, 19, 415-428.
- Lin, L. (2018). Bias caused by sampling error in meta-analysis with small sample sizes. *PLoS ONE* 13(9), e0204056. <https://doi.org/10.1371/journal.pone.0204056>.
- Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, 7(1), 105-125.

- Novianti, P. W., Roes, K. C., & van der Tweel, I. (2014). Estimation of between-trial variance in sequential meta-analyses: A simulation study. *Contemporary Clinical Trials, 37*, 129-138. DOI:10.1016/j.cct.2013.11.012.
- Nord, J. R. (1981). Three steps leading to listening fluency: A beginning. In H. Winitz (Ed.). *The comprehension approach to foreign language instruction*. Boston, MA: Newbury House.
- Oakes, J. (1990). Multiplying inequalities: *The effects of race, social class, and tracking on opportunities to learn mathematics and sciences*. Santa Monica, CA: Rand.
- Peterson, P. W. (1997). *Knowledge, skills, and attitudes in teacher preparation for content-based instruction*. New York: Longman.
- Polanin, J. R., & Pigott, T. D. (2013). The Campbell Collaboration's systematic review and meta-analysis online training videos. *Research on Social Work Practice, 23*(2), 229-232.
- R Core Team. (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available online at <https://www.R-project.org/>.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638-641.
- Schmidt, F. L., Oh, I., & Hayes, T. L. (2009). Fixed- versus random-effects models in meta-analysis: Model properties and an empirical comparison of differences in results. *British Journal of Mathematical and Statistical Psychology, 62*, 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.)*. Alexandria: ASCD.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1-48.
- Wilson, D. B. (2011). Effect size calculations and elementary meta-analysis. Retrieved September 9, 2019, from
- Wilson, R. M., & Ribovich, J. K. (1973). Ability grouping? Stop and reconsider. *Reading World, 13*, 84-91.

예시언어(Examples in): Korean

적용가능 언어(Applicable Language): Korean

적용가능 수준(Applicable Levels): Elementary/Secondary/College

박옥희

배재대학교(교수)

주시경대학 기초교육부

대전광역시 서구 배재로 155-40

35345

okpark@pcu.ac.kr

윤지여

정동초등학교(교사)

경상남도 사천시 정동면 대곡 1 길 25

52523

yjymodu@empas.com

Received: October 16, 2019

Reviewed: November 27, 2019

Revised version received: December 5, 2019