creative commons

COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

교육학 석사 학위논문

학습곤란 학생을 위한 데이터기반교수(DBI)의 효과크기 분석: 단일대상설계 연구를 중심으로

2021년 08월

서울대학교 대학원

협동과정 특수교육전공

최 서 현

학습곤란 학생을 위한 데이터기반교수(DBI)의 효과크기 분석: 단일대상설계 연구를 중심으로

지도교수 김 동 일

이 논문을 교육학 석사 학위논문으로 제출함
2021년 6월

서울대학교 대학원
협동과정 특수교육전공
최 서 현

최서현의 석사 학위논문을 인준함
2021년 7월

위 원 장 _____

부위원장 _____

위 원 _____

국문초록

일반학교 교실에는 여러 가지 이유로 기초학습기능 습득 및 활용에 심한 어려움을 보이는 학생들이 있다. 여기에는 학습장애를 비롯하여 난독증, 정서 및 행동적 문제, 평균 이하 또는 경계선 수준의 지능, 다문화적 배경, 환경적 결손, 불충분한 교육 기회 등 다양한 요인의 영향이 존재한다. '학습곤란(learning difficulties)'이라 함은 읽기, 쓰기, 수학의 기초학습영역에서 현저히 낮은 성취와 어려움을 보여 추가적 지원이 요구되는 이질적인 학습자들을 가리키는 포괄적 용어이다. 교육현장에서 이들 내 하위집단을 명확히 구분하는 것은 상당히 어려운 일이며, 이 중에는 특수교육대상자로 진단되어 지원을 받는 학생들도 있지만 그렇지 못해 적절한 개입을 놓치는 경우도 많다.

이에 다양한 학습곤란 학생들의 어려움을 '특수한 교육적 요구(special educational needs)'로 인식하고, 진단 및 분류에 앞서 적절한 중재를 제공할 필요성이 제기되었다. 국내에서도 다양한 학습곤란을 지닌 학생들에 대한 관심이 점차 증대됨에 따라 각종 법적, 제도적 기반 마련 및 사업 수행 등 지원이 이루어져왔다. 그럼에도 불구하고 여전히 일반학급 내 학습곤란 학생의 수는 증가하고 있고 많은 수가 적절한 지원을 받지 못한 채 사각지대에 놓여 있는 것으로 보고되므로, 이들에게 효과적인 중재를 제공하여 실질적인 학습권을 보장하려는 노력이 적극적으로 이루어질 필요가 있다.

데이터기반교수(data-based instruction, DBI)은 학습곤란을 지니며 특히 일반적인 방식의 중재 강화에도 반응하지 않는 학생들에게 효과적이라고 보고되는 교수적 접근 방식이다. 이는 학생에게

연구에 기초한 중재를 제공하면서 진전도를 주기적으로 측정하고, 수집된 객관적인 데이터를 바탕으로 교수적 수정의 필요성과 그 시기, 방법 등을 결정하는 일련의 반복적인 과정을 일컫는다. 진전도 모니터링을 위하여 교육과정중심측정(curriculum-based measurement, CBM)에 기초한 측정도구가 사용되고, 데이터기반교수는 일반학교 전체학급에서 시작하여 소집단, 개별중재로 나아가는 다층 체제에서 적용될 수 있다.

한편 특정한 중재 및 교수전략이 학생들에게 활발하게 적용되기 위해서는 그것이 과학적으로 입증되어 증거가 확보된 증거기반실제(evidence-based practice)임을 밝혀야 한다. 이를 위해서는 첫째, 관련 연구들이 체계적인 연구 설계 및 방법으로 수행되었는지에 대한 질적 분석, 둘째, 연구결과에서 밝히는 효과의 크기가 일정 수준 이상인지에 대한 확인이 요구된다. 국내에서는 교육과정중심측정 원리에 따른 기초학습능력 수행평가체제(Basic Academic Skills Assessment, BASA)가 개발된 이후 데이터기반교수 연구가 꾸준히 진행되어 왔음에도 불구하고, 그것이 증거기반실제로서 갖춘 수준은 어떠한지 확인한 연구는 부족한 실정이다. 따라서 본 연구에서는 국내에서 학령기 학습곤란 학생을 대상으로 데이터기반교수를 적용한 연구들을 종합하여 질적 분석과 메타분석을 통한 효과크기 분석을 실시하고, 이를 통해 증거기반실제를 위한 자료를 제공하고자 진행되었다.

본 연구에서 활용된 분석대상논문은 단일대상설계 연구 32편이고, 크게 두 가지 연구방법을 활용하여 분석하였다. 첫째, 데이터기반교수 연구의 연구설계 및 방법 측면의 질을 확인하기 위하여 Horner 외(2005)의 질적 지표에 기초하여 질적 분석을 실시하였다. 둘째, 학령기 학습곤란 학생에게 데이터기반교수가 미치는 효과를 양적으로 종합하기 위하여 메타분석을 실시하였다. Van den

Noortgate & Onghena (2003a; 2003b; 2007)에서 제시한 방법을 활용하여 데이터 표준화 및 효과크기 계산을 한 후 HLM 프로그램을 활용하여 2수준 다층메타분석을 실시하였다. 이를 통해 연구참여자 수준의 전체 효과크기를 산출하고, 효과크기의 이질성을 검증하였으며, 효과크기의 이질성에 영향을 주는 예측변인을 탐색하였다.

본 연구의 주요 결과는 다음과 같다. 첫째, 국내 학령기 학습곤란 학생 대상으로 데이터기반교수 연구는 2000년 이후 지속적으로 이루어져온 것으로 확인되었다. 그러나 상당수의 연구에서 어떠한 기준 및 절차에 따라 교수적 수정, 데이터기반 의사결정을 하였는지를 구체적으로 제시하지 않고 있었다. 둘째, 질적 지표를 활용한 분석 결과, 데이터기반교수 연구는 전반적으로 충분한 질적 수준을 갖춘 것으로 분석되었다. 다만 중재제공자, 중재충실도, 측정신뢰도, 실험통제 등 연구결과의 신뢰도를 높이기 위하여 더 나은 질적 수준 확보가 필요한 측면도 있는 것으로 나타났다. 셋째, 다층메타분석 결과 연구참여자 수준에서 데이터기반교수의 표준화된 전체 효과크기는 1.34로 상당히 높은 수준으로 나타났고, 통계적으로 유의했다. 넷째, 데이터기반교수의 효과는 연구참여자에 따라 이질적인 것으로 나타났다. 예측변인이 효과에 미치는 영향을 확인한 결과, 학생유형 중 지적장애의 유무, 전체 회기 수, 1회기 당 중재 시간이 효과크기의 이질성을 유의하게 설명하는 것으로 나타났다. 다섯째, 교사에 대한 지원은 데이터기반교수의 효과를 유의하게 높이는 변인으로 보고되는 반면, 국내에서 실시된 데이터기반교수에 있어 교사 지원이 이루어진 경우는 거의 없는 것으로 나타났다. 본 연구에서는 이러한 분석 결과에 따라 논의, 제한점 및 후속 연구에 대한 제언을 제시하였다.

주요어 : 데이터기반교수(DBI), 학습곤란, 질적 지표, 메타분석, 증거기반실제
학 번 : 2019-29606

목 차

Ⅰ. 서론 ····· 1
1. 연구의 목적 및 필요성 ····· 1
2. 연구 문제 ····· 5
3. 용어의 정의 ····· 5
1) 학습곤란을 지닌 학생 ····· 5
2) 데이터기반교수 ····· 6
3) 증거기반실제 ····· 6

Ⅱ. 이론적 배경 ····· 7
1. 학습곤란을 지닌 학생 ····· 7
1) 학습곤란의 개념 ····· 7
2) 다양한 학습곤란을 지닌 학생 ····· 8
3) '특수한 교육적 요구'로서 학습곤란 ····· 11
4) 학습곤란을 지닌 학생을 위한 지원 ····· 14
2. 데이터기반교수 ····· 15
1) 데이터기반교수의 개념 및 효과 ····· 15
2) 데이터기반교수의 역사적 배경 ····· 18
3) 교육과정중심측정 ····· 20
4) 기초학습영역에서의 CBM 측정 유형 ····· 22
5) 데이터기반 의사결정 ····· 24
3. 기초학습기능 수행평가체제 ····· 25
1) 기초학습기능 수행평가체제 개관 ····· 25
2) 데이터기반교수를 위한 BASA 검사의 활용 ····· 29
3) BASA를 적용한 국내 연구 동향 분석 ····· 30
4. 특수교육연구에서의 증거기반실제 ····· 35

1) 질적 지표 ····· 35
2) 메타분석을 통한 효과크기 산출 ····· 37
5. 학습곤란 학생 대상
데이터기반교수의 선행연구 고찰 ····· 38
1) 미국의 선행연구 고찰 ····· 38
2) 국내 선행연구 고찰 ····· 40
3) 국내 (준)집단실험설계 연구의
질적 수준 및 효과크기 분석 ····· 41

Ⅲ. 연구방법 ····· 50
1. 자료 수집 ····· 50
2. 포함 및 배제준거 ····· 51
3. 데이터 코딩 ····· 54
1) 질적 지표 분석을 위한 데이터 코딩 ····· 54
2) 메타분석을 위한 데이터 코딩 ····· 55
4. 데이터 분석 ····· 58

Ⅳ. 연구결과 ····· 62
1. 분석대상논문의 특성 ····· 62
2. 질적 지표 분석 결과 ····· 70
3. 효과크기 분석 결과 ····· 75
1) 모형 설정 ····· 75
2) 동질성 검증 및 평균 효과크기 산출 ····· 77
3) 기초학습영역별 평균 효과크기 산출 ····· 79
4) 예측변인의 조절효과 분석 결과 ····· 81

Ⅴ. 논의 및 결론 ····· 85

1. 논의 및 결론 ····· 85
2. 연구의 제한점 및 제언 ····· 96

참고문헌 ····· 100
부록 ····· 123
Abstract ····· 134
Summary ····· 138

표 목 차

<표 Ⅱ-1> 출판연도에 따른 분석 결과 ····· 31
<표 Ⅱ-2> 학술지명에 따른 분석 결과 ····· 32
<표 Ⅱ-3> 연구주제에 따른 분석 결과 ····· 34
<표 Ⅱ-4> Gersten 외(2005)와 Horner 외(2005)의
질적 지표 ····· 36
<표 Ⅱ-5> 집단설계 연구의 주요 특징 ····· 46
<표 Ⅱ-6> 집단설계 연구의 데이터기반교수 관련 특징
및 효과크기 ····· 47
<표 Ⅱ-7> 집단설계 연구의 질적 지표 분석 결과 ····· 49
<표 Ⅲ-1> 메타분석을 위한 코딩틀 ····· 56
<표 Ⅳ-1> 단일대상설계 연구의 개요 ····· 62
<표 Ⅳ-2> 단일대상설계 연구의 일반적인 특징 ····· 64
<표 Ⅳ-3> 단일대상설계 연구의 연구참여자 관련 특징 ·· 65
<표 Ⅳ-4> 단일대상설계 연구의 중재 관련 특징 ····· 66
<표 Ⅳ-5> 단일대상설계 연구의
데이터기반교수 관련 특징 ····· 69
<표 Ⅳ-6> 단일대상설계 연구의 질적 지표 분석 결과 ···· 73
<표 Ⅳ-7> 단계별 분석 모형 ····· 77
<표 Ⅳ-8> 무조건 모형 분석 결과 ····· 78
<표 Ⅳ-9> 기초학습영역별 무조건 모형 분석 결과 ····· 80
<표 Ⅳ-10> 조절효과 분석에 투입된 예측변인 ····· 81
<표 Ⅳ-11> 학생 관련 예측변인 분석 결과 ····· 83
<표 Ⅳ-12> 중재 관련 예측변인 분석 결과 ····· 84
<표 Ⅳ-13> 데이터기반교수 관련 예측변인 분석 결과 ···· 84

그 림 목 차

[그림 Ⅱ-1] 데이터기반교수의 절차 ····· 17
[그림 Ⅲ-1] PRISMA 4단계 도식 ····· 54
[그림 Ⅳ-1] 연구참여자별 효과크기에 대한 시각적 분석 · 79

Ⅰ. 서 론

1. 연구의 목적 및 필요성

학교 교실에는 다양한 특수한 교육적 요구를 지닌 학생들이 존재한다. 그 중에서도 기본적인 학습기능 습득 및 활용에 심각한 어려움을 경험하는 학생들이 있다. 현재의 교육 시스템에서 이러한 학습에 심각한 곤란을 지닌 학습자들이 자신의 교육적 요구에 적합한 집중적인 특수교육지원을 받기 위해서는 특수교육법의 장애 영역 내 진단을 받아야 한다(최수미 외, 2018). 이들 중에는 학습장애로 진단을 받은 학생도 포함되어 있으나, 엄격한 진단 기준을 충족하지 못했거나, 낙인에 대한 두려움으로 진단을 거부하거나, 장애로 진단되기까지 오랜 시간을 기다려야 하거나, 기타 이유로 방치되어 있는 등 다양한 이유로 일반학교현장에서 적절한 교육적 지원을 받지 못하고 있는 경우가 있다(김동일, 고혜정, 2018). 이에 진단 및 명칭 부여에 앞서 이들이 가지는 특수한 교육적 요구에 집중하여, 조기에 개별화되고 효과적인 교육적 서비스를 제공할 필요성이 제기되고 있다. 이는 특수교육의 패러다임이 장애, 진단에서 지원, 서비스, 교육으로 전환됨을 반영하는 것이기도 하다(손승현, 이예다나, 2017; 강종구, 김라경, 2012).

기초학습기능에 심각한 어려움을 겪는 학생들의 특성은 매우 다양하고, 실무적으로 이들 내 하위 집단을 명확히 구분하는 데는 어려움이 따른다. 실제로 많은 학습장애 학생들은 학습부진 집단에 속해 있을 것으로 추정된다(이대식, 2019). 이에 이들을 보다 포괄적인 개념으로 칭하고, 구체적인 '분류 및 진단'에 앞서 '지원'이 우선되어야 한다는 인식이 제기되어 왔다. '학습곤란(learning difficulties)'을 지닌 학생은 다양한 원인으로 인하여 학업 영역에서 낮은 성취나 학습에의 어려움을 보이는 이질적인 집단을 일컫고(Graham & Bailey, 2007), '학습 어려움' 또는 넓은 범

위의 '학습부진' 등으로 불리기도 한다. 특별한 교육적 지원이 필요한 학생들을 구분하고 일컫기 위해 학습장애(learning disabilities), 교육사각지대 학습자(at-risk learners), 저성취(low achievement), 학습부진(under-achievement) 등 다양한 개념들이 있는데, 학습곤란은 이러한 개념들을 아우르는 포괄적인 용어라고 볼 수 있다. 이는 결과로서 나타나는 심각한 학습 문제에 어느 수준 이상의 영향을 가지는 복합적인 요인들이 존재할 수 있음을 인정한다(김동일, 고혜정, 2018). 이들이 특수교육 적격성 단계에 의뢰되기 전에 교사와 학교 차원에서 이들의 다양한 학습곤란을 해소하기 위해 적절한 교육적 지원을 제공할 필요성이 제기되어 왔다(Richards et al., 2007). 이는 최근 많은 국가에서 특수교육대상자를 '특별한 요구를 가진 학생(students with special needs)'의 개념으로 보고 있으며, 특수교육지원의 전체적인 접근 방식이 일반학급 내 교육적 지원으로 변화하고 있는 움직임을 반영한다(손승현, 이예다나, 2017).

국내의 경우 다양한 학습곤란을 지닌 학생들에 대한 관심이 증가됨에 따라 법적, 제도적 기반이 마련되고 이들을 지원하기 위한 여러 사업들이 수행되어 왔다. 2015년 「초·중등 교육법 개정안」발의와 2016년 「초·중등 교육법」제28조(학습부진아 등에 대한 교육) 개정을 통해 특수교육대상자는 아니지만 특별한 요구가 있는 학생 및 '느린 학습자'를 위한 지원에 대한 법적 근거가 마련되었고, 「장애인 등에 대한 특수교육법」제15조에 경계선 지능을 가진 학습자에 대한 내용이 포함되었으며, 2017년 「기초학력 보장법안」을 통해 5년을 주기로 기초학력보장 종합계획을 수립하도록 하였다(이애진, 이예다나, 2019; 최수미 외, 2018). 이 외에도 두드림학교, 학습도움센터 및 클리닉 센터, 교육복지사업, 기초학력책임지도제, 경계선상학교 지원, 학력향상형 창의경영학교 사업, 맞춤형 진단·보정 자료 활용 및 기초학력 진단보정 시스템, 방과후 보충학습지도, 대학생 및 외부 강사에 의한 멘토링, 방학 중 학습지도 특별 프로그램 등 학습곤란 학생들을 위한 학습 및 교육, 정서 및 행동문제, 심리상담, 특수교육, 교육복지 등 다양한 측면에서 학습곤란 학생들을 위한 지원이 이뤄져왔다(이대식, 2015; 2019). 그러나 이러한 노력

에도 불구하고, 여전히 일반학급 내에서 다양한 학업적 어려움을 지닌 학생들의 수가 지속적으로 증가하고 있고(이대식, 임건순, 2019), 이들 중 많은 수가 적절한 선별 및 지원을 받지 못한 채 방치되고 있는 것으로 보고된다(신재현, 2019a). 이들에 대해 교사 및 학교의 효과적인 개입이 적시에 제공되지 않는다면, 학습곤란 학생들의 학습 실패는 누적되고, 이차적인 정서, 행동적 문제로도 이어질 가능성이 높다(김동일, 고혜정, 2018).

학습곤란을 지닌 학생들을 위한 교육적 지원은 충분한 이론적 기반 및 경험적 근거에 기초하고 있을 때 효과적이다. 데이터기반교수(data-based instruction, DBI)는 학습장애를 비롯하여 심각한 학습 문제를 지닌 학생들에게 있어 '증거(evidence)'에 기반한 효과적인 교수 방법이라고 제시된다(National Center on Intensive Intervention, 2013; Deno & Mirkin, 1977). 데이터기반교수는 기초학습능력 또는 행동적 영역에서 학생들의 데이터에 기초하여 언제, 그리고 어떻게 중재를 수정 및 강화할 것인지 결정하는 일련의 반복적이고 지속적인 과정을 의미한다(NCII, 2013). 일반적으로 기초학습영역에서 낮은 성취를 보이는 학생들을 위해 교사의 전문성 신장, 교수 집단의 크기 축소, 추가적인 행동적 지원 제공, 중재 전달 관련 요인들에의 변화 등 여러 가지 접근이 사용되는데, 이러한 접근에도 반응하지 않고 여전히 기대되는 성취 및 진전을 보이지 못하는 학생들이 있다(Fuchs, Fuchs, & Malone, 2017; Vaughn & Fuchs, 2003). 이러한 학생들에게 있어 데이터기반교수는 높은 효과를 가지고, 읽기, 철자/쓰기, 수학 영역 성취를 유의미하게 증진시킨다고 보고된다(Stecker, Fuchs, & Fuchs, 2005).

국내에서 데이터기반교수의 개념은 특히 한국형 교육과정중심측정(curriculum-based measurement, CBM) 도구인 기초학습능력 수행평가체제(Basic Academic Skills Assessment, BASA)가 개발된 2000년을 기점으로, 학습곤란을 보이는 학령기 학생들을 효과적으로 가르치는 목적으로 도입, 적용되어 왔다(김동일, 2000). 이후 학습곤란 학생을 위해 데이터기반교수를 적용하려는 연구 및 현장의 노력이 꾸준히 전개되어 왔

다. 그러나 아직까지는 국내에서 이들을 위해 데이터기반교수가 어떻게 실시되어 왔으며, 기존에 수행된 연구들이 증거로 수용가능한지 검증한 연구가 부족한 실정이다.

과학적 연구와 교육 실제 간 연계가 강조됨에 따라 교육현장에서는 과학적으로 입증되어 증거가 확보된 중재를 제공할 것이 권고된다(Odom et al., 2005). 최근 특수교육 연구 분야에서도 증거기반실제(evidence-based practice)를 확인하고 증거기반 중재의 기준을 확립하고자 하는 노력이 진행되고 있다(이예다나 외, 2015). 어떤 중재 및 교수전략이 증거기반실제로서 효과적임을 밝히고 그에 따라 교육 현장에서의 적용을 권고하기 위해서는 관련 연구들이 체계적인 연구 설계 및 방법으로 수행되었으며 중재에 따른 효과의 크기가 일정 수준 이상임이 입증되어야 한다(Odom et al., 2005). 데이터기반교수 자체에서도 성공적인 효과를 이끌기 위해서는 그 질과 충실도는 필수적인 조건이라고 제시된다(NCII, 2013). 이에 국내에서 실시되어 온 데이터기반교수 역시 그것이 충분한 증거에 기초한 효과적인 접근인지 확인하고, 현장에 더욱 활발히 적용하기 위해서 경험적 근거가 확인될 필요가 있다.

따라서 본 연구에서는 국내 학령기 학습곤란 학생을 대상으로 수행된 데이터기반교수 연구를 종합하여 질적 분석 및 메타분석을 실시하고 그 효과를 파악하고자 한다. 구체적으로 질적 지표를 활용하여 연구설계 측면에서의 질적 수준을 확인하고, 메타분석을 통해 여러 선행연구들을 종합한 데이터기반교수의 효과크기를 도출하고자 한다. 또한 국외 선행연구에 따르면 데이터기반교수의 효과크기는 사용된 CBM 측정도구, 교사지원, 교수적 수정 등 여러 변인에 따라 차이가 있다고 보고된 바 있기에(Filderman et al., 2018; Jung et al., 2018), 국내에서 수행된 데이터기반교수의 효과에 영향을 미치는 주요 변수들을 확인하는 데도 목적이 있다. 국내 선행연구에서는 학령기 학습곤란 학생들을 대상으로 데이터기반교수의 효과를 종합하여 구체적으로 살펴보지 않았는데, 예컨대 여승수, 홍성두, 손지영(2014)에서 국내 CBM의 연구동향을 종합하였으나 CBM의 측정학적 적합성에 초점을 두고 있고, 메타분석에 의한 연구 종

합이 아닌 서술적 종합에 해당되었다. 정평강(2018)의 연구에서는 증거기반개별화(data-based individualization)를 중심으로 기존의 문헌을 종합하였으나, 문헌 선정 기준에 따라 최종적으로는 해외 연구들만이 분석에 포함되었다. 따라서 본 연구는 데이터기반교수를 주제로 실시된 기존의 연구들을 종합하여 연구설계 측면의 질을 검토하고 메타분석을 실시한다는 점에서 선행연구들과의 차이점을 가지고, 증거로서의 수준을 밝히는 기초자료로 활용될 수 있다.

2. 연구 문제

본 연구는 국내에서 기초학습영역에서 어려움을 보이는 학령기 학습곤란 학생을 대상으로 수행된 데이터기반교수 연구를 종합하여 효과를 확인하고자 한다. 이를 위한 구체적인 연구 문제는 다음과 같다.

1) 학령기 학습곤란 학생들을 대상으로 기초학습영역에서 데이터기반교수를 제공한 연구들의 특성 및 연구의 질은 어떠한가?

2) 학령기 학습곤란 학생들을 대상으로 기초학습영역에서 수행된 데이터기반교수의 평균 효과크기는 어떠한가?

3) 학령기 학습곤란 학생 대상 데이터기반교수의 효과크기에 영향을 미치는 변인(학생 관련 변인, 중재 관련 변인, 데이터기반교수 관련 변인)으로는 무엇이 있으며, 그 영향은 어떠한가?

3. 용어의 정의

1) 학습곤란을 지닌 학생(students with learning difficulties)

학습곤란을 지닌 학생은 기초학습기능 습득 및 활용에 심한 어려움을 보여 추가적인 교육적 지원이 요구되는 다양한 학생들을 가리키는 포괄적인 용어이다(van Kraayenoord & Elkins, 2004; 김동일, 고은영, 2018). 본 연구에서는 학습장애, 학습부진, 저성취, 난독증, 경계선 지능, 행동 및 정서적 어려움, 다문화적 배경, 환경적 결손 등 다양한 요인들로 인해 기초학습영역인 읽기, 쓰기, 수학에서 심각한 어려움을 보이는 학생들을 학습곤란을 지닌 학생, 또는 학습곤란 학생이라고 부르고자 한다.

2) 데이터기반교수(data-based instruction, DBI)

데이터기반 개별화(data-based individualization), 데이터기반 프로그램 수정(data-based program modification), 증거기반교수 등으로 불리는 데이터기반교수(data-based instruction)는 효과적인 중재를 제공함과 동시에 신뢰로운 측정 도구로 학생의 진전을 지속적으로 점검하며, 수집된 학생의 성취 데이터를 활용하여 중재의 강화 및 수정을 결정하는 일련의 반복적인 과정을 일컫는다(Fuchs, Fuchs, & Stecker, 2010; NCII, 2013). 이는 특정 교수 프로그램 및 전략이기보다는 선정된 프로그램 및 전략을 적용함에 있어 적용되는 교수적 접근이다. 데이터기반교수에서 진전도를 측정하기 위해 교육과정중심측정(CBM) 원리에 따라 개발된 도구가 활용되고, 이는 학습장애 진단 및 선별을 위한 중재반응모형(responsiveness-to-intervention, RTI)이나 핀란드형 3단계 학습지원모델과 같이 '일반학급 전체-소집단-개별'의 다층체계(multi-tiered framework) 내에서 실시될 수 있다.

3) 증거기반실제(evidence-based practice)

증거기반실제 또는 근거기반실제는 과학에 기반을 둔 중재라는 의미로, 미국 특수아동협회(Council for Exceptional Children)에서는 인과관

계를 도출할 수 있는 연구설계를 바탕으로 학생의 성과에 유의미한 효과가 있음을 밝히는 양질의 연구가 충분히 있는 실제로 정의하였다(정광조, 이대식, 2014). 국내 특수교육 현장에서 증거기반실제는 “어떠한 특성 및 특별한 교육적 요구를 가진 학생에게 어떠한 특성의 중재 및 교수활동을 적용하였더니 학생에게 어떠한 성과가 있다고 연구결과에서 밝혀져 있으므로, 현재 관심대상 학생의 특성, 학습환경, 수행준거 등을 고려할 때 이러한 중재 및 교수활동을 제공하는 것이 적절하다는 과학적 의사결정 과정”으로 정의된 바 있다(최종근, 2010). 어떠한 중재가 증거기반실제로 고려될 수 있는지를 확인하기 위하여 질적 분석과 메타분석이 활용될 수 있다. 각각의 중재가 체계적인 연구방법 및 설계에 따라 실시되었는지 확인하기 위해(Odom et al., 2005), 연구의 질을 검증하기 위한 질적 지표들이 제시되어 활용된다. 또 두 개 이상의 개별 연구결과들을 종합하여 단일한 효과크기를 제시하기 위해(Higgins et al., 2019), 메타분석이 활용된다.

Ⅱ. 이론적 배경

1. 학습곤란을 지닌 학생

1) 학습곤란의 개념

‘학습곤란(learning difficulties)’은 기초적인 학습에 심각한 어려움을 보여 학업을 위한 추가적 지원이 요구되는 광범위한 특성의 학생들을 가리킨다(van Kraayenoord & Elkins, 2004). 여기에는 난독증, 행동 및 정서적 장애, 다문화적 배경, 가정에서의 관심 및 지원 결손 등 복잡한 요인들로 인해 현저히 낮은 학업 성취를 보이는 이질적인 학습자들이 포함

된다(Elkins, 2000; 김동일, 고혜정, 2018). 이대식(2019)에서는 많은 학습장애 학생들이 학습부진 집단에 포함되어 있을 것으로 추정된다는 점에서 학습부진 및 학습장애 학생을 지칭하여 '학습 어려움 학생(students with learning difficulties)'라고 칭하였다. 또 김애화 외(2018)에서는 「장애인 등에 대한 특수교육법」의 '학습장애', 「난독증학생 지원 조례」의 '난독증', 「초·중등 교육법」의 '학습부진', 「기초학력보장법안」의 '학습지원대상학생' 개념을 비교한 결과, '학습지원대상학생'의 개념이 내적 원인의 유무와 관계 없이 학업 성취의 심각한 어려움을 보일 경우 선정될 수 있는 것으로 가장 포괄적인 개념이라고 보았다. 본 연구의 학습곤란을 지닌 학생은 이러한 학습지원대상학생의 의미에 가깝다.

2) 다양한 학습곤란을 지닌 학생

학습곤란 안에는 학습부진(under-achievement), 저성취(low achievement), 학습장애(learning disabilities), 경계선 지능(borderline intellectual functioning), 난독증(dyslexia) 등이 모두 포함되는데, 각각의 개념을 구분하여 제시하면 다음과 같다. 정의에 명확한 합의는 없으나 일반적으로 학습부진은 학생의 IQ 점수를 바탕으로 기대되는 성취에 비해 실제 학업성취가 현저하게 낮은 경우를 의미하고, 저성취는 개인의 지적 능력과는 관계 없이 또래 학생들의 성취 수준에 비해 현저히 낮거나(하위 5% 또는 하위 20%) 평균 수준보다 낮은 경우를 의미한다(김동일, 이대식, 신종호, 2016). 한편 국내 「초·중등 교육법」에서는 학습부진을 '성격장애나 지적 기능의 저하 등으로 인하여 학습에 제약을 받는 학생 중 「장애인 등에 대한 특수교육법」 제15조에 따른 학습장애를 지닌 특수교육대상자로 선정되지 아니한 학생'으로 정의하고 있다. 이때의 학습부진에는 경계선지능을 지닌 학습자가 해당되어(김애화 외, 2018), 용어가 쓰이는 맥락에 따라 지칭하는 집단에 다소 차이가 있을 수 있다.

이와 달리 학습장애는 앞선 학습부진 및 저성취와 같이 잠재적 원인들로 인해 나타나는 '증상'으로서의 낮은 성취가 아닌 신경인지적 특성에

기인하는 실질적인 진단명이다(김동일, 이대식, 신종호, 2016). DSM-5에서는 특정학습장애(specific learning disorder)를 신경발달장애의 한 유형으로 보는데, 읽기, 쓰기, 수학 등 학업적 기술을 배우거나 사용하는 데에서의 어려움을 의미한다(American Psychiatric Association, 2013). 구체적으로 '부정확하거나 느리고 부자연스러운 읽기', '읽은 것의 의미를 이해하는 것의 어려움', '맞춤법의 어려움', '쓰기 표현의 어려움', '수 감각, 수에 관한 사실, 산술적 계산 숙달에서의 어려움', '수학적 추론에서의 어려움' 중 한 가지 이상의 증상을 6개월 이상 나타낼 경우 진단된다(APA, 2013). 우리나라의 「장애인 등에 대한 특수교육법」에서는 학습장애를 '개인 내적 요인으로 인하여 듣기, 말하기, 주의집중, 지각, 기억, 문제해결 등의 학습기능이나 읽기, 쓰기, 수학 등 학업성취 영역에서 현저하게 어려움이 있는 사람'으로 정의하고 있다(김동일, 이대식, 신종호, 2016; 김애화 외, 2018). 연구 및 사용되는 맥락에 따라 정의에 다소 차이가 존재하나, 일반적으로 학습장애라 함은 '개인 내적인 신경생물학적 특성에서 기인하고, 학습 상의 문제는 학교에서의 수행 및 일상생활을 현저하게 방해하는 수준이어야 하며, 발달적, 신경학적, 감각 또는 운동의 손상, 정신적 장애나 환경적, 문화적, 경제적 결손 등으로 인한 것이 아니어야 함'을 공통적인 요소로 포함하고 있다(APA, 2013; IDEA, 2004; Lyon et al., 2001).

한편 우리나라에서 특수교육대상자 중 학습장애로 진단받은 학생의 비율은 꾸준히 감소하고 있는데, 2020년 기준 전체 특수교육대상자 중 1.3%에 불과한 것으로 나타난다(국립특수교육원, 2020). 이는 학습장애가 전체 특수교육대상자 중 30% 이상으로 가장 높은 비율을 차지하는 미국과는 상반되는 결과이다(강은영, 김우리, 신재현, 2020). 그런데 학습장애 비율의 감소 추세와 달리 우리나라에서 기초학력을 미달하는 학생의 비율은 지속적으로 증가하는 추세인데(교육부, 2019), 이는 학습장애를 지닌 학생 중 상당수가 제대로 진단되지 못해 적절한 지원을 제공받지 못하고 있음을 암시한다고 볼 수 있다. 아직까지도 국내에서 학습장애의 개념 정립이 모호하여 학습부진, 경도지적장애 등과 명확한 구분이 어렵

고, 실질적인 선별과 진단을 위한 절차에 대한 현장의 합의와 이해가 부족하여 학습장애로 진단되어야 하는 학생들이 진단을 받지 못하고 있는 실정이다(김동일 외, 2012; 민수진, 2019). 실제로 선행연구에 따르면 학교 현장에서 상당 수의 학습장애 학생은 저성취 학생, 느린 학습자, 기초학습부진 학생 또는 다른 유형의 장애를 가진 학생으로 분류되어 왔을 가능성이 있고, 그들의 학습 문제가 훨씬 더 심각해진 후에야 판별될 수도 있다(김동일, 이대식, 신종호, 2016; 민수진, 2019). 이 외에도 특수교육대상자의 낙인 문제로 인해 진단을 거부하거나 특수교육지원을 받기 위해 장기간을 기다려야 하는 등의 문제로 인해 교육적 지원이 시급함에도 불구하고 방치되어 있는 경우가 존재한다(최수미 외, 2018).

다음으로 경계선 지능과 난독증 학생 또한 심각한 학습 곤란에도 불구하고 그 어려움이 제대로 인식되지 못하거나 적절한 지원을 받지 못하고 있는 학생 집단에 해당된다. 경계선 지능은 일반적으로 평균에서 -1 표준편차 이하인 71부터 84까지의 지능지수로 정의되는데, 이들은 지적장애 범주에 포함되지는 않으나 일반학생에 비해 인지적 어려움을 지닌다. 난독증은 읽기장애의 하위유형에 해당되는데, 단어인지의 정확성 및 유창성의 어려움과 철자 및 해독의 어려움을 보이는 것을 의미한다(Lyon, Shaywitz, & Shaywitz, 2003). 경계선 지능, 난독증을 비롯하여 다양한 요구를 가진 학습자에 대한 관심이 증대됨에 따라 이들을 위한 정책적 지원 및 관련 연구가 이뤄지고 있음에도 불구하고, 이들은 여전히 합의된 정의의 부족, 선별 및 진단 도구의 부족, 법적, 제도적 기반의 미비 등으로 인해 사각지대에 놓여있다(강옥려, 2016; 이애진, 이예다나, 2019). 특히 경계선 지능 학생들은 학령기 초기부터 학업, 인지, 사회성, 정서 및 행동 등의 여러 측면에서 어려움을 경험하고 많은 경우 또래로부터의 고립 및 따돌림의 대상이 되며(한국보건복지인력개발원 아동자립지원단, 2017; Chauhan, 2011), 학령기부터 시작된 학습 및 관계의 누적적 실패는 이후에도 정신과적 문제, 범죄, 직업적 어려움, 가정생활의 어려움 등으로 이어질 가능성이 높다고 보고된다(Hassiotis et al., 2008; 한국보건복지인력개발원 아동자립지원단, 2017).

일반학교 교실에는 상당히 많은 수의 경계선 지능 및 난독증 학생들이 존재하는 것으로 추정되고 있고, 많은 경우 이들은 일반교육과 특수교육 사이에 위치하여 학습곤란에 대한 적절한 중재를 받지 못하는 실정이다. 지능의 정규분포를 가정할 때 전 인구의 13.6% 정도가 경계선 지능 범주에 해당되며, 국내에는 한 학급에 세 명꼴로 전국적으로 80만 명의 경계선 지능 학생이 존재할 것으로 추정된다(변관석, 신지숙, 2017; 최수미 외, 2018). 또 2015년 교육부에서 제시한 '난독증 현황파악 연구보고서'에 따르면 전국 154개 초등학교 초등학생 8,575명 중 1%가 난독증, 2.2%가 난독증 고위험군, 1.4%가 난독증 저위험군으로 추정되었다(이애진, 이예다나, 2019). 이들 외에도 사회문화적 배경, 교육적 기회 결손, 정서 및 행동적 문제 등 다양한 요인으로 학습곤란을 겪는 학생들의 비율이 상당할 것이며, 이들에 대한 교육적 지원이 시급하다.

3) '특수한 교육적 요구'로서 학습곤란

국제적으로 특수교육의 패러다임은 장애에 대한 의학적 관점에서 사회적 관점으로 전환되고, '장애'에서 '지원'으로 초점이 변해왔다(손승현, 이예다나, 2017; 강종구, 김라경, 2012). 예컨대 2001년 세계보건위원회에서 승인되어 대한민국을 포함한 30개국 이상에서 사용되고 있는 국제 기능장애와 건강분류(International Classification of Functioning, Disability, and Health, ICF)에서는 '기능과 장애'와 '상황적 요소'의 두 개 영역과 그 안의 '신체 기능 및 구조', '활동과 참여', '환경적 요소들', '개별적 요소들' 네 가지 구성요소를 통해 장애를 설명한다(WHO, 2001). 이는 장애에 대한 환경 변인을 더욱 강조하게 되었으며 장애를 개인과 환경 간 상호작용 맥락 속에서 이해하고, 장애에 대한 환경 변인을 더욱 강조하며, 나아가 지원의 범주를 확대하려는 시도를 반영한다(홍성두 외, 2019; McDougall, Wright, & Rosenbaum, 2010).

이렇듯 장애에 대한 관점의 변화와 함께 특수교육대상 여부를 결정함에 있어서도 장애 진단 및 능력 여부에 초점을 두는 방식에서 지원과 연

결되는 요구에 기준을 두는 방식으로 변화해야 함이 강조되고 있다(손승현, 이예다나, 2017). 즉 장애 자체에 집중하기보다는 학교 교육과정, 학습환경 등 학생이 놓인 교육적 맥락과 개별 특성을 연결시켜 구체적인 어려움과 요구를 확인하고 그에 적절한 중재를 제공할 필요가 있다는 것이다(허승준, 2016). 실제로 장애에 대한 진단은 서비스 제공의 토대가 되는 것은 사실이나, 진단명이 학생의 독특하고 개별적인 요구를 확인하는 데 가림막이 되기도 하고, 학습장애와 같은 경우에는 개념 정의 및 진단 평가 상의 혼란으로 인해 지원 대상에 포함되지 않고 사각지대로 밀려나는 등 한계가 있다고 지적된다(손승현, 이예다나, 2017).

이러한 맥락에서 OECD(2003)가 제시한 특수교육이 필요한 아동에 대한 범국가적 범주는 학습곤란의 의미를 '특수한 교육적 요구'로서 이해하는 데 도움이 된다. OECD의 분류 체계에 따르면 '특수한 교육적 요구를 지닌 학생들(students with special educational needs)'은 세 가지 범주를 포괄하는데, A범주의 '장애(Disabilities)', B범주의 '곤란/어려움(Difficulties)', C범주의 '불리(Disadvantages)'가 그것이다. 먼저 '장애(Disabilities)'의 범주에는 합의된 진단이 가능한 신체기관의 병리로 인한 장애 또는 손상이 있는 사람이 포함되고, 이들의 교육적 요구는 주로 이러한 장애에 기인하는 것으로 여겨진다(Robson, 2005). 다음으로 '곤란/어려움(Difficulties)'의 범주에는 행동 및 정서적 장애 또는 학습장애가 있는 사람이 포함되는데, 이들의 교육적 요구는 주로 학생과 교육적 맥락 간 상호작용에서 발생한 문제에 기인하는 것으로 여겨지고(Robson, 2005), 이들의 어려움이 학교 학습에서 기인한 것인지, 내적 문제로 인한 것인지, 또는 사회적 불리로 인한 것인지는 명확하지 않다(김지연, 2017a). 마지막으로 '불리(Disadvantages)'는 주로 사회경제적, 문화적 또는 언어적 측면에서 발생하고, 주로 이러한 요인으로 인한 불리에서 이들의 교육적 요구가 기인한다고 본다(Robson, 2005). 이러한 세 범주로의 분류는 학생의 교육적 어려움의 지각된 원인에 기초하고 있다(Robson, 2005).

이러한 OECD의 '특수한 교육적 요구를 지닌 학생들'에 대한 분류는

장애에 대한 사회적 모델을 반영하여 신체적, 인지적 장애부터 사회경제적 결손까지 복합적인 이유로 학습 상 어려움을 겪는 학생들을 포괄한다는 장점이 있다(OECD/European Communities, 2009). 더불어 이는 학생의 특수한 교육적 요구와 관련하여 추가적인 교육적 지원이 제공된다는 점이 전제로 되어야 함을 강조하는 자원기반 접근방식(resource-based approach)에 해당되어(Robson, 2005), 다양한 학습곤란 학생들에 대한 지원을 가능케 한다. OECD의 분류체계는 범국가적이나 국가에 따라 특수교육에서 다루는 범위 및 범주 내 명칭은 상이하다. 이 중 영국, 덴마크, 네덜란드의 경우 '비범주적 체계'를 도입하여 모든 학생의 개별화된 요구에 대한 지원 필요성을 인정하고, 장애와 학습곤란을 모두 가진 학생, 장애만 가진 학생, 장애가 없이 학습곤란만 가진 학생 모두를 지원 대상으로 포함하고 있다(손승현, 이예다나, 2017). 우리나라의 경우 A범주의 '장애(Disabilities)' 안에 시각장애, 청각장애, 지적장애, 지체장애, 자폐성장애, 건강장애, 발달지체를, B범주의 '곤란/어려움(Difficulties)'에 정서·행동장애, 학습장애가 포함되는 것으로 보고 이를 특수교육에서 담당하고 있다(김지연, 2017a). 2020년 기준 B범주에 해당되어 지원을 받는 정서·행동장애, 학습장애 학생은 각각 1,993명, 1,226명으로, 전체 특수교육대상자 중 차지하는 비율은 약 3.4%에 해당되는 인원이다(교육부, 2020). 이는 특수교육 지원을 받는 전체 학생 중 B 범주에 해당되는 정서혼란(emotional disturbance)과 특정학습장애(specific learning disabilities) 학생의 비율이 2017년 기준 43.7%에 달하는 미국에 비해 현저히 낮은 수치이다(U. S. Department of Education, 2019).

본 연구에서 다루는 학습곤란을 지닌 학생은 대부분 OECD 분류체계에 따라 B범주인 '곤란/어려움(Difficulties)'에 해당되는 것으로 볼 수 있다. 즉 학습장애로 진단된 학생을 비롯하여 특수교육대상에 해당되지는 않으나 심각한 학습의 어려움으로 인한 특수교육적 요구가 있고 이에 개입이 필요한 학습부진, 저성취 등의 학생을 포괄한다. 다만 본 연구에서는 적은 연구 수이지만 경도 수준의 지적장애 학생, 경계선지능 학생 또한 포함되어 있다는 점에서 A와 B 범주를 아우른다고 볼 수 있다.

4) 학습곤란을 지닌 학생을 위한 지원

학습곤란을 지닌 학생들 중에는 학습장애 등 특수교육대상자로 진단되어 교육적 지원을 받는 학생들도 있지만 심각한 학업적 어려움에도 불구하고 아직 진단되지 못해 적절한 개입과 중재를 놓치는 경우가 많다. 국내에서는 교육 현장에 적절한 교육을 받지 못한 채 방치되는 다양한 학습곤란 학생들이 존재함을 인식하고 이들을 실질적인 학습권을 보장하기 위한 현장 및 학계의 노력이 진행되어 왔다. 특히 최근에는 특수교육대상자로 분류되지는 않았지만 특별한 교육적 요구를 지닌 학생들에 대한 관심이 증가되고 있고, 이들을 지원하기 위한 정책적 노력과 관련 연구들이 지속적으로 이뤄지고 있다(이애진, 이예다나, 2019).

특히 2010년부터 국내에서는 학습장애 진단 및 선별에 있어 미국의 중재반응모형(RTI)을 도입하여 활용하고 있는데, 이는 연구에 기초한 양질의 효과적인 중재를 일정 기간 제공했음에도 불구하고 기대되는 성취 및 진전을 보이지 않는 경우에 학습장애로 진단하는 접근법을 말한다(Bender & Shores, 2007). 이는 중재반응모형 전까지 주된 학습장애 진단 및 선별 모형이었던 능력-성취 불일치 모형(ability-achievement discrepancy model)에 비해 몇 가지 장점을 가진다. 능력-성취 불일치 모형의 경우, 개인의 지적 능력(IQ)에 비해 현저히 낮은 학업 성취를 보이는 학생을 학습장애로 선별하는데, 학업 실패를 보일 때까지 기다린 뒤 교육적 서비스를 제공한다는 점, 지적능력과 학업성취 간 완벽에 가까운 상관을 가정한다는 점, 중재 계획에 정보를 제공하지 않는다는 점 등의 제한점이 있었다(여승수, 홍승두, 2011). 이에 비해 중재반응모형에 기초한 학습장애 진단 및 선별에서는 학습의 어려움이 심화되기 전 조기 선별 및 개입이 가능하고, 선별 과정에서부터 교육적 지원이 제공되며, 단 한 번의 검사 결과에 의존하지 않는 등의 장점이 있다(김동일, 이대식, 신종호, 2016). 이러한 접근에서는 중재에 대한 반응을 확인하기 위해 전체 학급 학생들 또는 학습에 어려움을 보이는 학생들을 대상으로

연구에 기초한 효과적인 중재를 선제적으로 제공하게 된다. 이로써 진단을 받은 학생에게만 중재가 제공되는 것이 아니라 이를 통해 진단 전 개입에 더불어 적격성을 충족하지 못하거나 배제요인에 해당되어 특수교육 대상자인 학습장애로 선정되지 않은 위험학생들에게도 지원을 제공할 수 있게 된다. 최근에는 미국의 중재반응모형에서 더 나아가 일반교육과 특수교육의 협력, 학생의 교육적 요구에 따른 다양한 지원 형태, 적극적인 조기 중재가 더욱 강조되는 핀란드의 3단계 교육지원체제에 대한 논의도 이루어지고 있다(김애화 외, 2018; 이대식, 2019).

또한 국내에서 학습곤란 학생들을 지원하기 위해 다양한 사업이 수행되어 왔다. 각 교육청에서 수행되고 있는 사업들을 정리하면, 두드림학교, 학습맘, 방과 후 보충학습지도, 방학 중 학습지도 특별프로그램, 학습전략 및 학습 상담, 학습도움센터 및 클리닉 센터 상담 지원, 정서·심리지원을 위한 위클래스, 마을 공동체, 맞춤형 진단·보정 자료 활용 및 기초학력 진단보정시스템, 대학생 등에 의한 멘토링, 집중 캠프, 교육복지사업, 지역사회 연계, 학생 맞춤형 기초·기본학력 향상 프로그램, 기초학력책임지도제 등이 있다(이대식, 2019). 그러나 여전히 많은 수의 학습곤란 학생이 일반학교에서 적절한 지원을 받지 못한 채 사각지대에 놓여있다고 보고되고(신재현, 2019a), 학습에서의 실패가 계속 누적될 경우 자신에 대한 부적절감, 무기력함 등 심리적 어려움과 이차적인 정서, 행동적 문제로도 이어질 가능성이 높으므로 더욱 관심이 필요다(김동일, 고혜정, 2018).

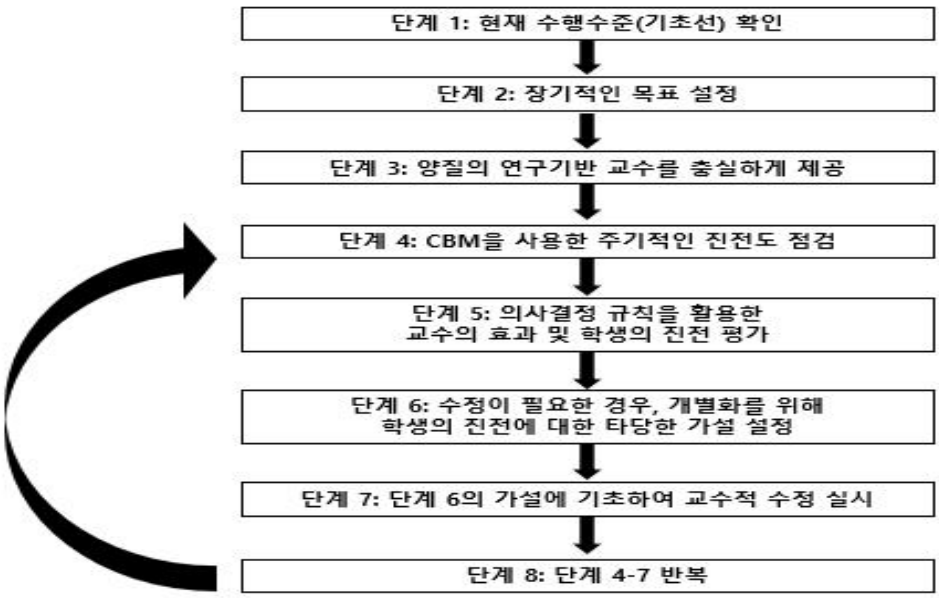
2. 데이터기반교수

1) 데이터기반교수의 개념 및 효과

특수교육 분야에서는 학습 상의 심각한 어려움을 겪는 학생들을 위해 전문적이고 효과적인 교수적 접근들을 개발, 평가, 적용하려는 노력이 꾸

준히 있어왔다(Jung et al., 2018). ‘데이터기반교수(data-based instruction, DBI)’은 이러한 교수적 접근의 하나로, 이는 ‘평가 데이터를 활용하여 읽기, 수학, 또는 행동에 대한 개입을 강화 및 수정하는 시기와 방법을 결정하는 체계적 방법’으로 정의된다(NCII, 2013). 이는 효과적 중재를 제공한 뒤 중재에 대한 반응을 나타내는 학생 성취 데이터를 활용하여 교수적 수정을 하는 방법이다(Fuchs, Fuchs, & Stecker, 2010). 데이터기반교수는 ′데이터기반 개별화(data-based individualization)’(NCII, 2013), ‘데이터기반 프로그램 수정(data-based program modification)’(Deno & Mirkin, 1977), ‘실험적 교수(experimental teaching)’(Casey et al., 1988) 등으로도 불린다.

선행연구에 따르면 데이터기반교수의 개념은 다음의 절차를 공통적으로 포함한다(Jung et al., 2018; Lembke et al., 2018). ‘(1)학생의 현재 성취 수준 확인하기, (2)장기적인 교육 목표 세우기, (3)양질의 증거기반 중재(evidence-based intervention)를 충실하게 제공하기, (4)학생의 진전을 빈번하게 모니터링하기, (5)교수적 수정의 필요 여부를 결정하기 위해 데이터기반 의사결정 규칙(data-based decision rules)을 활용하기, (6)학생의 구체적인 요구에 대한 잠정적인 가설을 수립하고, 가설에 기초하여 교수적 변화를 가하기, (7)진전도를 모니터링하며 데이터를 바탕으로 교수적 수정의 효과를 평가하기, (8)이러한 절차들을 학생이 학업 목표를 달성할 때까지 반복하기’가 그것이다. 이를 그림으로 표현하면 [그림 Ⅱ-1]과 같다. 교수적 수정 및 강화를 결정하고 시행하기 위해서는 학생의 진전도를 주기적으로 모니터링할 것과 개별 학생의 데이터에 기초할 것이 요구된다(Jung et al., 2018).



[그림 Ⅱ-1] 데이터기반교수의 절차 (Lembke et al., 2018)

데이터기반교수는 특수교육 서비스 차원에서 장애학생을 대상으로 할 수도 있고, 또는 학습장애 선별 및 진단을 위한 중재반응모형을 비롯하여 일반교육 현장에서 학습 및 행동 지원을 위한 다층체계 체제 내에서 실시될 수도 있다(Fuchs, Fuchs, & Compton, 2012; Dennis & Gratton-Fisher, 2020). 미국 증거기반중재 연구소(American Institutes for Research Center on Response to Intervention)에서 제시한 RTI 모델에 따르면, RTI는 선별(screening), 다층의 교수(multi-level instruction), 진전도 점검(progress monitoring), 데이터기반 의사결정(data-based decision-making)의 네 가지 요소로 구성되어 데이터기반교수와 맥을 같이 함을 알 수 있다(Oslund, Elleman, & Wallace, 2020).

학습곤란 학생을 위한 데이터기반교수의 효과성은 여러 연구에서 지속적으로 보고되어 왔고(Stecker, Fuchs, & Fuchs, 2005; Fuchs, Fuchs, & Hamlett, 1989; Poch, McMaster, & Lembke, 2020), 비단 학생 뿐 아니라 교사들로 하여금 교수전략을 계획하고 수정 및 평가하는 데도 도움이 되는 것으로 제시된다. 연구에 따르면 교사들은 데이터기반교수를 할 때 더욱 구체적인 교수 계획을 세웠고, 교수적 조정을 더욱 빈번하게 하

였으며, 학생들에게 적합하고 보다 현실적인 목표 기술을 설정할 수 있었다고 보고된다(Capizzi & Fuchs, 2005; Fuchs, Deno, & Mirkin, 1984).

2) 데이터기반교수의 역사적 배경

1975년 모든 장애아동을 위한 교육법(Education for All Handicapped Children Act, PL 94-142)이 제정됨에 따라 특수교육분야에서 교수와 평가가 연계될 것이 의무화되었고, 개별화교육(Individualized Education Programs, IEP)에 목표에 대한 진전도 평가 절차를 포함시키도록 강조되었다. 이에 특수교사들은 교수적 프로그램을 평가, 개선하기 위하여 개별 학생의 진전도를 지속적으로 모니터링할 필요가 있었다(Deno & Mirkin, 1977). 이에 더하여 관찰에 기초한 평가의 부정확성에 대한 문제점이 지적되면서 객관적인 측정 및 평가와 교수 간의 연계가 강조되었다(Fuchs, Deno, & Mirkin, 1984).

이러한 흐름 속에서 반복적이고 체계적인 평가 시스템인 '데이터기반 프로그램 수정(data-based program modification)'(Deno & Mirkin, 1977) 개념이 등장하였다. 데이터기반 프로그램 수정은 교사들로 하여금 일련의 단기 목표 대신 장기 목표에 대한 학생의 진전을 측정하도록 하였고, 교사가 직접 개발한 평가 자료가 아닌 심리측정학적 적합성을 갖춘 검사와 엄격한 가이드라인에 입각하여 측정하도록 하였으며, 수집된 데이터를 활용하는 데 있어 구체적인 기준을 제시하였다는 점에서 특징적이다(Fuchs, Deno, & Mirkin, 1984).

한편 1980년대부터는 미국 미니에폴리스 공립학교(Minneapolis Public Schools, MPS)를 중심으로 학습장애 및 경도장애 영역에서 표준화된 규준 지향 검사로 교육적 의사결정을 내리는 것이 많은 제한점을 지닌다고 지적되기 시작하였다(Marston et al., 2003). 전통적인 의사결정 방법이 지니는 대표적인 문제점으로는 지능검사가 지니는 한계, 비영어권 학습자를 비롯한 다양한 배경의 학생들에 대한 평가의 편파성, 검사 실시를 위한 학교 심리학자들의 지나친 시간 소모, 평가와 교수 간 비연계성 등

이 있었다(김동일, 1998; Marston et al., 2003). 이에 미니에폴리스 공립학교(MPS)에서는 학습장애 및 경도장애를 위한 교육적 의사결정의 대안적 방법으로 문제해결모델(problem solving model, PSM)을 제안하였고, 이는 미니에폴리스 모델(Minneapolis model)이라고 불리기도 하였다. 특히 미네소타주 교육위원회(Minnesota State Board of Education)에서 문제해결모델을 전통적인 적격성 판정 기준에 대한 대안적인 방법으로서 승인함에 따라 1992년 9월부터는 미니에폴리스 공립학교(MPS)에서 이러한 데이터 기반 의사결정 접근(data-based decision making approach)을 공식적으로 사용할 수 있게 되었다(Marston et al., 2003).

문제해결모델에서는 지능검사 점수에 의존했던 전통적 기준을 넘어서 중재에 대한 학생의 반응을 바탕으로 적격성 평가를 실시하고, 이는 특수교육 분야의 측정 및 평가에 있어 획기적인 변화로 평가된다(Marston et al., 2003). 문제해결모델에서는 학생들의 학업적 어려움을 발견, 지원하기 위해 데이터에 기반한 체계적 절차를 활용하고, 일반학급에서부터 특수교육까지의 협력적 팀 접근이 강조된다. 구체적으로 문제해결 절차를 살펴보면, '(1)문제 확인 및 기술, (2)중재를 위한 전략의 수립 및 실시, (3)학생의 진전도 모니터링 및 중재의 효과성 평가, (4)필요 시 절차의 반복'으로 요약할 수 있다(Marston et al., 2003). 특히 교육과정중심측정(curriculum-based measurement, CBM)이라는 대안적 평가방법은 전통적인 규준 지향 검사에 비해 더욱 민감하고 유용한 정보를 제공하기에, 문제해결절차의 체계적 적용에 있어 핵심적인 기능을 했다(김동일, 1998).

이렇듯 문제해결모델에서는 교수와 평가가 밀접한 연계를 이루어 반복되고('teach-test-teach-test'), 이를 통해 교수에 대한 학생의 반응에 입각하여 학생별 최소 제한 환경(least restrictive environment)을 계획할 수 있다(Marston et al., 2003; Jenkins, Deno, & Mirkin, 1979). 또한 미니에폴리스 공립학교(MPS)의 문제해결모델은 '전체 학생을 대상으로 하는 일반학급에서의 중재(단계 1), 다학문적 팀의 협력 하 학생의 어려움 및 해소 방안에 대한 가설을 수립하는 단계(단계 2), 집중적인 중재에

도 충분한 반응을 보이지 않는 학생들에 대하여 특수교육 적격성을 평가하는 단계(단계 3)'의 연속적인 세 단계로 구분할 수 있다(Marston et al., 2003). 이렇듯 특수교육에서의 교수 및 평가를 지속적인 문제해결 절차로 이해하는 관점은 이후 중재반응모형(response to intervention, RTI), 데이터기반교수(DBI) 등의 개념으로 발전하였다.

3) 교육과정중심측정

데이터기반교수에서는 계속적인 교육적 의사결정을 내리기 위하여 단기간에 학습곤란 학생의 성장을 민감하게 인지할 수 있는 신뢰로운 검사도구의 사용이 중요하다(Jung et al., 2018). 1970년대 미네소타 대학에서 Stanley Deno와 그의 동료들에 의해 개발된 교육과정중심측정(curriculum-based measurement, CBM)은 이러한 목적에 부합하는 대안적인 평가방법이다. CBM은 학생의 기초학습능력 수준과 진전을 쉽고 간편하면서도 정확하게 측정하기 위한 표준화된 검사로(Deno, 1985), 학생의 수행수준을 반복적으로 점검하고, 그 데이터를 활용하여 현재 제공되고 있는 교수의 효과를 평가하여 효과적이지 않을 경우 목표, 전략 등을 수정하기 위한 측정도구이다(Stecker, Fuchs, & Fuchs, 2005). 개발 당시 특수교사들의 교육적 의사결정 과정을 지원하기 위해 신뢰롭고 타당한 데이터를 제공하려는 목적에서 출발하였다(Deno, 1985). 데이터기반교수 절차에서 CBM 도구는 수집된 데이터를 바탕으로 교수적 목표를 상향조정할지, 교수 내용 및 방법을 그대로 유지할지, 변화를 가할지 등을 결정하는 데 활용된다(McMaster et al., 2020).

일반적으로 학생의 수행 수준을 확인하기 위한 평가 방법에는 여러 가지 유형이 존재하는데, 그 중 학생이 직접 배우는 교육과정과 연계하여 수행 수준을 점검하는 방법을 교육과정중심사정 또는 교육과정중심평가(curriculum-based assessment, CBA)라고 일컫는다(Hintze, Christ, & Methe, 2006). CBM은 CBA의 한 유형으로, 현재 학교에서 사용되는 교육과정에 입각하여 수행능력을 측정함으로써 교육적 의사결정에 필요한

정보를 얻고자 하는 방법이다(김동일, 1998). 일반적으로 CBA는 무엇을 측정하는가에 따라 두 가지 유형으로 구분될 수 있다. 첫째는 현재 교육과정에서 다루는 특정 기술에 대한 학생의 도달 여부를 측정하는 것(specific subskill mastery measurement)이고, 둘째는 일반적인 성취도를 측정하는 것인데(general outcome measurement), 이 중 CBM은 후자에 속한다(이승미 외, 2018; 여승수, 2010). 또 CBA의 대부분은 비형식적이고 표준화되지 않은 절차를 이용하는 것에 비해, CBM은 규준, 표준화된 측정 절차, 전산화된 관리 체계를 포함한다는 점에서 구분된다(김동일, 1998).

CBM은 실시가 간편하고 시간이 적게 드는 특징 덕분에 지역 규준을 만들기 쉽고, 집단의 특성을 적절히 반영하는 지역 규준을 만드는 것은 특수교육에서 적절한 의사결정을 돕는다(김동일, 1998). CBM 점수는 해당 학습 영역에 대한 또래 집단의 점수 분포를 바탕으로 산출된 개별 학생의 일반적인 능력 수준을 대표한다고 해석된다(Deno et al., 2001). 즉 교사는 CBM 점수를 학생의 일반적 능력 수준에 대한 지표로 여길 수 있으며, 또래 집단의 성취를 바탕으로 개인 간 차이를 식별할 수 있다. 이는 “학생이 보이는 성취 수준과 성장 정도가 학급, 학교, 또는 지역구의 다른 학생들에 비해 얼마나 차이가 있는가?”에 대한 한 가지 답을 제시한다(Deno et al., 2001). 특히 학교 및 지역구마다 교육과정이 다른 미국에서 학교의 수준, 목표 등에 따라 달라질 수 있는 교육과정 내의 특정 기술을 측정하기보다 CBM을 통해 읽기, 수학, 쓰기 등 기초학습 영역에서의 일반적인 성취 수준을 측정하고자 한 점은 의의가 크다. 이러한 특성 덕분에 CBM은 교육과정이 없는 단위학교에서도 적용될 수 있다(Deno, 2003). 또한 학습곤란을 지닌 학생들의 경우 그 성취와 진전도가 또래 학생들에 비해 낮은 편이므로, 일반적으로 사용되는 기초학습능력 측정도구로는 그들이 알고 있는 것과 성장하는 정도를 민감하게 측정하기 어렵다. 이에 비해 CBM은 학습곤란 학생들의 성장에 민감하고, 현저히 낮은 성취 및 진전을 바탕으로 개별 학생의 특수한 교육적 요구를 확인하는 데 도움이 되며, 이를 바탕으로 현재 학습장애 위험군 학생들

의 보편적 선별(universal screening)의 일부로서 널리 사용되고 있다(Fuchs, 2004).

한편 어떠한 측정도구가 학교 현장에서 실질적으로 활발히 활용되기 위해서는 여러 조건들이 요구된다. CBM은 이와 관련한 강점들을 지니고 있는데, 특히 뛰어난 효율성 및 간편성과 높은 측정학적 적합성을 특징으로 한다(여승수, 홍성두, 손지영, 2014; Jung et al., 2018). 즉 실시에 드는 비용이 낮고, 실시 및 채점, 해석이 수월하며, 성장에 민감하면서도 측정학적 적합성이 높다(Romig, Therrien, & Lloyd, 2017). 이러한 조건은 실제 학교 현장에서 검사도구가 활용되기 위해 중요하게 고려되는 점이다. 만약 검사 실시가 불편하고 많은 시간과 노력이 필요하다면 아무리 강점이 많다고 하더라도 교사들이 활발하게 사용하기 어려울 것이다. CBM은 일반적으로 실시하는 데 1분에서 5분이 소요되고, 이는 검사를 위해 따로 특정 시간을 지정할 필요 없이 수업 전이나 학교 일과 시간 도중에 간편하게 실시할 수 있음을 뜻한다(여승수, 홍성두, 손지영, 2014). 효율성에 더하여, 높은 측정학적 적합성, 즉 신뢰도와 타당도 또한 검사의 활용에 있어 매우 중요하다(여승수, 홍성두, 손지영, 2014). CBM의 경우 그 개발 과정에서 전통적인 심리측정학의 주요 개념과 행동 및 관찰 측정론을 통합하였기에 다른 측정도구들에 비해 신뢰도 및 타당도에 더 큰 강조점을 두고 있다(Deno et al., 2001). 또한 CBM은 동일한 구인을 측정하는 동등한 난이도의 동형검사를 활용하여 반복 측정함으로써 상대적으로 넓은 범위의 학습 능력 및 기술에 대한 정보를 수집한다(Deno et al., 2001). 이를 통해 학생의 성취와 성장에 대해 보다 정확하고 의미 있는 정보를 제공할 수 있게 된다.

4) 기초학습영역에서의 CBM 측정 유형

기초학습영역인 읽기, 쓰기, 수학에서 주로 활용되는 CBM 검사 유형을 살펴보면 다음과 같다. 먼저 읽기 영역에서는 주로 구두읽기 또는 읽기유창성(oral reading/reading fluency) 검사와 빈칸 채우기(maze) 검사

의 두 가지가 활용된다. 읽기유창성 검사에서는 학생들이 주어진 지문을 소리 내어 1분 동안 읽고, 이 중 맞게 읽은 음절 수 또는 낱말 수를 기록한다. 빈칸 채우기 검사에서는 일곱 번째마다의 낱말이 삭제 또는 다른 세 개의 선택지로 대체되어 있는 지문을 주로 2분에서 3분 동안 음독 또는 묵독으로 읽고, 그 중 학생들이 맞게 고른 낱말 수가 기록된다. 이러한 CBM 읽기 검사는 표준화 읽기 성취검사와도 높은 상관을 보이며, 일반화된 읽기 능력에 대한 타당한 지표로 여겨진다(Nese et al., 2011; Wayman et al., 2007). 다만 빈칸 채우기 검사의 경우 읽기유창성 검사에 비해 더 적은 관심을 받아왔는데, 연구에 따르면 빈칸 채우기 검사 결과는 단어 읽기 및 읽기이해 능력에 대해 신뢰로운 정보를 제공하여 일반적인 읽기능력을 나타내는 지표로 고려할 수 있다(Kendeou, Papadopoulos, & Spanoudis, 2012).

쓰기 능력의 측정은 다른 영역에 비해 더 어려운데, 이는 쓰기 능력이 듣기, 말하기, 읽기 능력의 습득 이후에 발달하여 복합적인 특성을 지니며 단순한 글자쓰기에서부터 자신의 생각을 글로 표현하는 것까지 그 수준의 범위가 넓기 때문이다(김동일, 2002; Romig, Therrien, & Lloyd, 2017). CBM 쓰기 검사의 경우 주로 짧은 시간동안 단어 또는 문장 따라쓰기검사(word/sentence copying), 단어 받아쓰기 검사(word dictation), 그림-단어 검사(picture word), 서두 제시형 이야기 검사(story prompt) 등을 통해 쓰기 샘플을 수집하고, 계량적인 방식으로 쓰기 능력을 측정하게 된다(신재현, 2019b). 채점방식 또한 다양한데, 쓴 전체 단어 수(Words Written, WW), 철자를 바르게 쓴 단어 수(Words Spelled Correctly, WSC), 연속적으로 바르게 쓴 단어 수(Correct Word Sequences, CWS), 연속적으로 바르게 쓴 단어 수에서 틀린 순서를 제외한 단어 수(Correct Minus Incorrect Word Sequences, CIWS) 등이 연구 및 현장에서 주로 사용되어 왔다(Romig, Therrien, & Lloyd, 2017). 읽기에서와 마찬가지로 CBM 쓰기 검사 또한 일반적인 쓰기 능력 수준을 예측하는 의미 있는 데이터를 제공한다고 제시된다(Deno, 1985).

CBM 수학 검사는 2분 정도의 짧은 시간 동안 문제를 푸는 유창성

측정 형태로 실시된다(Clarke & Shinn, 2004). 기본 연산의 속도 및 정확성을 측정하는데, 이는 일반화된 수학 학습 능력에 대한 신뢰로운 지표로 알려져 있을 뿐 아니라 이후 고등 수준의 수학 능력 형성에도 결정적인 영향을 미친다고 보고된다(홍성두, 정은주, 김동일, 2006). 현재 교육 현장에서는 주로 구체적인 수업 내용에 대한 형성평가, 즉 숙달도 평가(mastery measurement) 방식으로 연산유창성 측정이 이루어지는데, 이는 구체적인 수업 내용에 대한 이해 정도를 확인하여 이후 교수 계획에 직접적으로 반영할 수 있다는 장점이 있지만 단기적인 교수목표 및 내용을 넘어 일반화된 수학 능력을 반영하지는 못한다는 한계도 있다(홍성두, 정은주, 김동일, 2006). 반면 CBM 수학 검사의 경우, 동형검사를 제작하기 위해서 일반적으로 연간교육과정에 기초하여 검사에 포함될 문제 유형 및 중요도를 확인하고, 중요도에 비례하여 각 유형의 문제들을 포함하는 절차를 거친다(Fuchs, Hamlett, & Stecker, 1991).

5) 데이터기반 의사결정

교사는 개별 학생의 진전도 데이터에 기초하여 타당한 교육적 의사결정을 내려야 한다. 일반적으로 데이터기반 의사결정(data-based decision making) 기준의 유형은 기울기 방법(slope method), 데이터포인트 방법(points below method/data point method), 숙달기준 방법(criterion mastery)의 세 가지가 제시되고 있다(Ardoin et al., 2013).

기울기 방법의 경우 학생의 성장을 나타내는 추세선(trend line)을 활용하고, 데이터포인트 방법의 경우 그래프 상의 연속적인 데이터포인트의 개수를 활용하여 학생의 목표선(goal line)과 비교한다(Jung et al., 2018). 추세선의 기울기가 목표선 아래로 떨어지거나, 특정 개수의 연속적인 데이터포인트가 목표선 아래로 떨어질 경우 교사는 교수적 수정을 하도록 권고되고, 추세선 또는 연속적인 데이터포인트가 목표선 위에 위치할 경우 목표를 상향조정할 수도 있다(Jugn et al., 2018). 국내 기초학습영역에서 활용되는 CBM 도구인 기초학습기능 수행평가체제(BASA)

에서는 기울기 방법과 데이터포인트 방법을 제시하는데, 특히 데이터포인트 방법의 경우 네 개 이상의 연속적인 점수가 목표선 아래로 떨어지거나 세 개 이상의 연속적인 점수가 목표선 위에 있을 때 조정을 하도록 권고한다(김동일, 2000; 2006; 2009). McMaster 외(2020)에서는 신뢰할 만한 진전도 기울기 값을 구하고 신뢰로운 의사결정을 하기 위해서는 교사가 최소 여덟 개 이상의 측정치를 확보하도록 권고하고 있다. 데이터기반교수 효과에 대한 메타분석을 실시한 Jung 외(2018)의 연구에서는 8개 데이터포인트 기준을 함께 활용할 때 효과크기가 가장 컸고, 4개 데이터포인트 기준과 추세선을 사용한 경우 또는 7개에서 10개 데이터포인트를 사용한 경우에도 긍정적인 효과가 있는 것으로 나타났다.

3. 기초학습기능 수행평가체제

1) 기초학습기능 수행평가체제 개관

우리나라의 경우 교육과정중심측정(CBM)의 개념은 1980년대 후반 특수교육분야에 도입되었고(Espin et al., 2012), 2000년에는 한국판 CBM인 기초학습기능 수행평가체제(Basic Academic Skills Assessment, BASA)(김동일, 2000)가 개발되었다. 이후 데이터기반교수 및 CBM 측정도구는 저성취 학생 지원을 중심으로 한국의 교육 분야에서 활발하게 적용되기 시작하였다(Espin et al., 2012). 우리나라의 CBM 연구는 학습장애 및 학습곤란과 같이 특수한 교육적 요구가 있는 학생들의 성장을 모니터링하는 시스템을 개발하는 데 초점이 있었고, 학교에서 실시한 교육적 프로그램의 효과를 평가하고 학생의 성장을 측정하기 위한 규준을 마련하고자 힘써왔다(Espin et al., 2012).

BASA(김동일, 2000)는 현재 국내에서 CBM 원리에 따라 기초학습영역에서 개발된 검사 중 상용화된 표준화 검사이다. 일반 및 특수교육에서 데이터기반교수를 제공할 때 주로 활용되며, 특히 학습장애 진단 및

학습장애 위험군 학생을 선별하거나 학생의 진전도를 점검하는 등 교육적 의사결정 과정에 활발하게 사용되고 있다. 특히 여승수, 홍성두, 손지영(2014)에서 밝히는 바와 같이 2003년을 기점으로 국내 CBM 연구가 활발해진 것은 BASA가 개발된 후 본격적으로 연구 및 교육 현장에 적용된 시점과 관련이 있다고 여겨진다.

BASA는 읽기, 쓰기, 수학, 수학문장제, 어휘, 읽기이해, 초기수학, 초기문해의 여덟 가지 기초학습영역에서 개발되었고, CBM 검사의 유형 및 원리를 반영하고 있다. 검사가 개발된 순서대로 각각을 살펴보면 다음과 같다.

첫째, BASA 읽기검사(김동일, 2000)는 초등학교 1학년에서 6학년을 대상으로 읽기 능력을 측정하며, 읽기 유창성과 빈칸 채우기 두 가지 세부검사로 구성된다. 읽기유창성 검사에서는 세 개의 동형검사를 바탕으로 1분 안에 학생들이 정확하게 읽은 음절수를 측정하고, 빈칸 채우기 검사에서는 지문을 3분 동안 묵독으로 읽으며 첫 문장을 제외한 모든 문장에서 일곱 번째 단어가 비어있어 세 개의 선택지 중 답을 고르게 된다(Kim et al., 2014). 읽기 유창성 검사의 동형검사 신뢰도는 .96이고, 빈칸채우기 검사의 Cronbach's α 계수는 .89이며, 기초학습기능검사(KEDI-Individual Basic Learning Skills Test)의 비형식적 독해력검사와의 상관이 .62~.80 범위에 있어 높은 측정학적 적합성을 보고한다(Kim et al., 2014).

둘째, BASA 수학(김동일, 2006)은 Ⅰ, Ⅱ, Ⅲ, 통합 단계로 구성되는데 이는 각각 초등학교 1학년, 2학년, 3학년, 1~3학년을 아우르는 수준에 해당되며, 학생의 현재 수준에 적합한 단계를 선택하여 실시한다(김자경, 강혜진, 김기주, 2016). 검사는 2분 동안 연산 유창성을 측정하는 문항들로 구성되어 있다. Spearman-Brown 공식으로 산출한 반분신뢰도 계수는 .73~.93이고, 맞은 자릿수에 따라 부분점수를 부여하는 CD(Correct Digits) 채점 방식을 사용하였을 때 ACCENT 수학검사와의 상관이 .54~.71로, 양호한 측정학적 적합성을 지님을 알 수 있다(김동일, 2006).

셋째, BASA 쓰기(김동일, 2009)는 이야기 서두검사(story starter) 제시형태로 실시되고, 3분 동안 정확하게 쓴 글자 수를 바탕으로 쓰기 유창성을 측정한다(김은향, 김동일, 고은영, 2013). 학생은 이야기 서두를 보고 1분 동안 무엇을 쓸지 생각한 후 3분 동안 작성한다. 정확 글자 수에 대한 정량적 평가에 더불어 오류 유형을 기록하고, 형식, 조직, 어휘, 표현, 내용, 주제에 대해 정성적 평가도 실시할 수 있다(김동일, 2009). BASA 쓰기 검사의 측정학적 적합성을 알아본 김동일, 김미순, 배성직(2003)에서 검사-재검사 신뢰도 및 동형검사 신뢰도는 .75 수준이고, 기초학습기능검사(KEDI-Individual Basic Learning Skills Test) 쓰기검사와의 상관은 평균 .55 이상으로 양호한 적합성을 지닌 것으로 나타났다.

넷째, BASA 초기문해(김동일, 2011a)는 만 4세 이상 아동을 대상으로 초기문해 수준을 측정한다. 기초평가는 음운처리과정을 평가하는 음운인식능력(phonological awareness), 음운적 작업기억능력(phonological working memory), 음운적 정보회상능력(retrieval of phonological information from long term memory)의 세 가지 소검사와 초기 읽기 능력을 평가하는 단어인지(word recognition)과 읽기유창성(reading fluency)의 두 가지 소검사로 구성된다. 형성평가는 이 중 음운인식능력 검사로 구성된다. 음운인식능력 검사는 음절과 음소를 변별, 합성, 탈락, 대치하는 과제로 이루어져있고, 음운적 작업기억능력은 숫자 회상 검사와 무의미 단어 회상 검사로 측정하며, 음운적 정보회상능력은 빠른 이름 대기 검사(rapid automatized naming test, RAN)으로 측정한다. 단어인지 검사의 경우 고빈도 단어, 저빈도 단어, 비단어를 활용하여 문자해독 능력을 측정하고, 읽기유창성 검사는 선택적으로 실시할 수 있다. 유치원 아동을 대상으로 측정한 BASA 초기문해의 검사-재검사 신뢰도는 각 소검사별로 .76~.97의 범위에 있었다(김희진, 2009; 김동일, 2011 재인용). 또한 공인타당도 검증을 위해 KISE-BAAT(Korean Institute for Special-Education Basic Academic Achievement Tests) 읽기 검사와의 상관을 소검사별로 살펴본 결과 .44~.65으로 양호한 수준으로 나타났다(김희진, 2009; 김동일, 2011 재인용). 만 4세반 아동 집단과 만 5세 아동

집단의 점수 차이를 확인한 결과, 음운적 작업기억 검사를 제외하고 모두 통계적으로 유의한 차이를 보여, 발달타당도 또한 적합한 것으로 나타났다(김희진, 2009; 김동일, 2011 재인용).

다섯째, BASA 초기수학(김동일, 2011b)은 만 4세 이상 아동을 대상으로 초기수학능력을 평가하는 검사로서, 수 인식(number identification), 빠진 수 찾기(missing number), 수량변별(quantity discrimination), 추정(estimation)의 네 가지 소검사로 구성된다. 수 인식 검사는 1부터 100까지의 수를 빠르고 정확하게 읽는 능력을 측정하고, 빠진 수 찾기는 1부터 20까지의 수 중 연속된 세 수에서 배열 규칙을 찾아 빠진 수를 인식하는 능력을 측정한다. 수량변별은 두 수 중 더 큰 수를 변별하는 능력을 측정하고, 추정 검사는 수직선 위에서 수의 위치를 추정하는 능력을 측정하며, 각 소검사는 1분 동안 실시된다. 초등학교 및 유치원 아동을 대상으로 BASA 초기수학의 측정학적 적합성을 확인한 결과, BASA 초기수학 검사 전체의 Cronbach's α는 .99로 매우 높았고, 모든 하위영역 및 모든 집단에서 Cronbach's α가 양호한 수준으로 나타났다. 공인타당도를 확인한 결과, KISE-BAAT 범자연수 영역과의 상관은 .87, K-WPPSI(Korean Wechsler Preschool and Primary Scale of Intelligence) 산수 소검사와의 상관은 .84, BASA 수학 검사와의 상관은 .78로 높은 수준으로 나타났다(허상, 2009; 김동일, 2011 재인용).

여섯째, BASA 수학문장제(김동일, 2018)는 언어적 능력과 계산적 능력을 복합적으로 요구하여, 수학의 최종 목표인 문제해결력을 측정하는 검사이다. 검사는 초등학교 3학년부터 6학년까지 총 4개 학년을 대상으로 하고, 약 10분 동안 실시된다. 검사는 학년별 국가수준 교육과정의 기본학습요소를 반영한 20문항으로 구성되고, 현행 교육과정에 입각한 대표 문항들로 이루어진다. 학생은 각 문항에 대하여 식과 답을 모두 작성해야 하고, 채점은 식 점수, 답 점수, 자릿수 점수를 각각 계산한다. 초등학교 3학년을 대상으로 검사의 신뢰도와 타당도를 확인한 김동일 외(2013)에서 BASA 수학문장제의 Cronbach's α는 .90으로 매우 높았고, BASA 수학과 BASA 읽기(빈칸채우기) 검사와의 상관이 각각 .44, .26으

로 나타나 양호한 측정학적 적합성을 지니고 있는 것으로 해석할 수 있다.

일곱째, BASA 어휘(김동일, 2019a)는 초등학교 3학년부터 6학년을 대상으로 현재 어휘력 수준을 측정하는 검사로, 15분 동안 실시되는 시간제한 검사이다. 3~6학년 국어 교과서에 제시된 어휘를 빈도분석 한 뒤, 고빈도 어휘, 저빈도 어휘, 공통어휘를 추출하고, 추출된 어휘를 동등한 비율로 활용하여 검사 문항이 개발되었다. 초등학교 3학년을 대상으로 BASA 어휘의 측정학적 적합성을 확인한 김동일 외(2016)에 따르면 전체 문항의 Cronbach's α는 .82로 매우 높게 나타났고, 검사-재검사 신뢰도는 .64로 양호한 수준이었다. 또한 확인적 요인분석 결과 CFI(comparative fit index), TLI(Tucker and Lewis index of fit), RMSEA(root mean square error of approximation) 지수의 값이 각각 .84, .81, .06으로 양호한 모형 적합도를 보였고, 하위 요인 간 상관관계가 .29~.56의 범위로 유의하여, 검사의 구인 타당도가 양호한 것으로 나타났다.

여덟째, BASA 읽기이해(김동일, 2019b)는 초등학교 3학년부터 6학년 학생의 읽기 이해력을 사실적 이해, 추론적 이해, 평가적 이해로 측정한다. 검사 문항은 오지선다형이며, 학년별 국가 수준 교육과정의 기본학습요소를 반영하는 대표문항들로 구성되어 있다. 검사는 15분 동안 실시된다. BASA 읽기이해의 Cronbach's α는 .87이고, 구인타당도 검증을 위한 확인적 요인분석 결과 CFI, TLI, RMSEA 값이 각각 .99, .98, .51로, 양호한 측정학적 적합성을 갖춘 것으로 볼 수 있다(김동일 외, 2020).

2) 데이터기반교수를 위한 BASA 검사의 활용

BASA는 CBM 원리를 따른 측정도구이자 측정학적 적합성이 높고 검사와 중재가 밀접하게 연결된다는 점에서 데이터기반교수에의 활용도가 매우 높다. BASA가 데이터기반교수 절차에서 진전도 데이터 수집, 교수의 개별화 등에 적용되는 방식은 다음과 같다(김동일, 2000; 2006; 2009).

먼저 BASA 검사 결과를 바탕으로 읽기, 쓰기, 수학의 현재 수행 수준을 확인하고 최초의 교수적 목표를 결정한다. 구체적으로는 세 번의 측정을 통해 기초선을 결정하고, 또래 규준에 비추어보았을 때 기대되는 월 진전도를 계산하여 목표선(target line)을 그리며, 이 목표선이 기본적인 교수적 목표가 된다(김동일, 2000; 2006; 2009).

다음으로 연구 증거에 기반한 양질의 교수를 제공함과 동시에, 학생의 성장을 모니터링하기 위해 주기적인 형성평가를 실시한다(김동일, 2000; 2006; 2009). BASA는 비슷한 난이도의 동형검사들로 이루어져 있어 진전도 측정에 적합하다. 또한 교수적 의사결정 규칙으로 제시되는 기울기 방법, 데이터포인트 방법, 숙달기준 방법 중 BASA에서는 기울기 방법과 데이터포인트 방법을 활용하도록 권고한다. 즉 데이터포인트 방법에 따라 네 개 이상의 연속적인 점수가 목표선 아래로 떨어지거나 세 개 이상의 연속적인 점수가 목표선 위에 있을 때, 또는 기울기 방법에 따라 성장선의 기울기와 목표선의 기울기 간 차이가 클 때 교사는 교수 목표, 내용, 방법 등에 변화를 가하게 되는 것이다.

3) BASA를 적용한 국내 연구 동향 분석

본 절에서는 2020년 12월까지 BASA를 적용한 국내 연구의 동향을 분석하여 제시하였다. BASA는 국내에서 CBM 원리에 기초하여 개발된 기초학습영역 측정도구 중 상용화된 표준화 검사이고, 현재까지 학습곤란을 지닌 학생을 위해 가장 빈번하게 적용되어 왔다. 이에 본 절에서는 BASA를 적용한 연구의 동향을 일차적으로 분석함으로써 학습곤란 학생 대상 데이터기반교수의 효과 및 질적 수준에 대해 본격적인 분석을 실시하기 전 '학습곤란을 지닌 학생', '교육과정중심측정', '데이터기반교수' 등과 관련하여 어떤 연구들이 수행되어 왔는지 확인하고자 하였다. 이에 기술적 내용분석방법(descriptive content analysis method)을 활용하여 BASA를 적용한 선행연구들의 주제를 파악하였다. 기술적 내용분석방법은 연구의 경향을 파악하는 데 목적이 있고(Cohen, Manion, &

Morrison, 2007), 특정한 주제를 중심으로 현존하는 데이터를 다시 분석 및 분류하는 방법이다(정동훈, 2009). 이 방법에서는 주제의 패턴을 드러내기 위해 주로 비율, 빈도 등 기술통계치를 활용한다(Dinçer, 2018).

기술적 내용분석 방법을 통해 BASA를 적용한 연구 동향을 개관하기 위하여 다음의 절차를 거쳤다. BASA를 적용한 논문을 수집하기 위하여 주요 논문검색사이트인 학술연구정보서비스(RISS), 한국학술정보(KISS), 누리미디어(DBPIA) 등에서 'BASA', 'Basic Academic Skills Assessment', '기초학습기능 수행평가체제', '기초학습기능' 등 주요 주제어를 조합하여 학술지 등재 논문과 석, 박사 학위논문을 검색하였다. 출판연도에는 제한을 두지 않고 2020년 12월까지 출판된 논문을 검색하였다. 누락을 최소화하기 위해 참고문헌 목록을 검토하여 추가하는 과정을 거쳤다. 본 절의 기술적 내용분석은 본 연구의 주제가 되는 메타분석 및 질적 분석과는 별개로 실시되었다. 다만 기술적 내용분석에 포함된 연구 중에는 이후 문헌 선정 기준을 충족하여 최종 분석에 포함된 경우도 있었다.

이상의 절차를 통해 137편의 학술지 및 학위논문이 선정되었다. 여기에는 다양한 목적으로 BASA 검사를 적용한 연구들이 포함되었는데, 예컨대 연구참여자 선정, 학업성취수준 측정, 검사도구의 측정학적 적합성 확인, 새로운 검사도구 개발 등의 다양한 목적으로 BASA가 활용되고 있었다. 출판연도, 학술지명, 연구주제의 세 가지 기준을 중심으로 기술적 내용분석 결과를 제시하면 다음과 같다.

<표 Ⅱ-1> 출판연도에 따른 분석 결과

출판연도	N	%
1999-2000	2	1.46
2001-2005	11	8.03
2006-2010	33	24.09
2011-2015	50	36.50
2016-2020	41	29.93
합계	137	100

<표 Ⅱ-1>에서 볼 수 있듯이 BASA와 관련된 연구들은 1999년도부터 시작되어 2015년까지 지속적으로 증가하는 경향을 보였다. 2011년부터 2015년 시기에 가장 많은 연구가 이루어졌는데(N = 50, 36.5%), 연구수의 증가 경향은 이후 지속되었다고 볼 수 없다.

<표 Ⅱ-2> 학술지명에 따른 분석 결과

분류	학술지명	N	%
사회과학			
	교육학		
	교과교육학연구	1	1.54
	교육심리연구	1	1.54
	교육종합연구(교육종합연구원)	1	1.54
	발달장애연구	3	4.62
	아시아교육연구	8	12.31
	언어치료연구	1	1.54
	열린교육연구	1	1.54
	유아교육연구	1	1.54
	정서, 행동장애연구	1	1.54
	통합교육연구	2	3.08
	특수교육	1	1.54
	특수교육교과교육연구	3	4.62
	특수교육연구	2	3.08
	특수교육저널: 이론과 실천	6	9.23
	특수교육학연구	6	9.23
	특수아동교육연구	5	7.69
	학교심리와 학습컨설팅	1	1.54
	학습자중심교과교육연구	1	1.54
	학습장애연구	7	10.77
	한국초등교육	3	4.62
	한국초등국어교육	1	1.54
	The SNU Journal of Education Research	1	1.54

기타사회과학		
특수교육재활과학연구	1	1.54
심리과학		
발달지원연구	1	1.54
한국심리학회지: 발달	1	1.54
사회과학일반		
육아정책연구	1	1.54
지적장애연구	2	3.08
자연과학		
생활과학		
아동학회지	2	3.08
합계 (학위논문: N = 72)	65	100

총 137편의 연구 중 학술지 논문에 해당된 연구는 65편(47.45%)이었다. 65편의 학술지 논문을 KCI 대분류와 중분류에 따라 분류한 결과는 <표 Ⅱ-2>와 같다. 대분류 체계에 따라 분류해보았을 때 대부분(N = 63, 96.92%)의 학술지 논문이 사회과학으로 분류되었고, 나머지 2편(3.08%)은 자연과학 분야에 해당되었다. 중분류 체계에 따라 보다 구체적으로 나누어보면, 높은 비율(N = 57, 87.73%)의 학술지 논문이 교육학 연구 분야였고, 기타사회과학, 사회과학 일반, 심리과학, 생활과학 연구 분야에서도 각각 한두 편씩 연구가 수행되었음을 확인하였다.

<표 Ⅱ-3> 연구주제에 따른 분석 결과

연구주제	N	%
학습부진	31	22.63
학습장애	25	18.25
지적장애 및 자폐스펙트럼 장애	22	16.06
CBM / 검사도구	11	8.03
일반교육	9	6.57
중재반응모형(RTI)	8	5.84
언어장애 및 언어치료	5	3.65
유아(학령전)	4	2.92
사회경제적지위(SES)	4	2.92
난독증	4	2.92
ADHD	3	2.19
학습곤란	3	2.19
다문화	2	1.46
부모	3	2.19
청각장애	1	0.73
저성취	1	0.73
장애학생 진로교육	1	0.73
합계	137	100

연구주제에 따른 분석 결과는 <표 Ⅱ-3>과 같다. 연구의 제목 및 초록에서 사용하고 있는 키워드를 중심으로 연구주제를 분류하였고, 한 개 이상의 키워드를 포함하고 있을 경우 연구자의 판단 하 더욱 초점이 된다고 여겨지는 주제를 중심으로 분석하였다. 분석 결과, 다양한 유형의 학습곤란이 주제로 다루어지고 있었는데, 그 중 학습부진을 중심으로 한 연구가 31편(22.63%)으로 가장 많았고, 학습장애를 다룬 연구가 25편(18.25%)으로 뒤를 이었다. 이 외에도 학습곤란과 난독증을 주제로 다루고 있는 연구도 각각 3편과 4편씩 있었다. 또한 BASA를 적용한 분석대상논문들 중에는 지적장애를 중심으로 한 경우도 21편으로 많았고, 일반

교육 분야에서도 9편으로 비교적 활발하게 연구되고 있었다. 이 외에도 CBM 및 검사도구, 유아기, 언어장애 및 언어치료, 중재반응모형(RTI) 등 비교적 다양한 연구주제를 중심으로 BASA가 적용되어 왔음을 확인할 수 있었다.

4. 특수교육연구에서의 증거기반실제

1) 질적 지표

교육 현장에서 교사들은 효과가 증명된 교수방법을 사용하여야 한다. 학생들은 연구에 기초한 표준화된 절차에 따라 충실도가 확보된 채로 중재를 받았을 때 성취가 증진될 수 있다(Fuchs et al., 2003). 현재 미국에서는 교육법에 의해 교사 및 연구자가 과학적으로 증명된 교육방식을 사용하도록 의무화하고 있다(홍이레, 2016). 국내외 특수교육계에서도 과학적인 방법 및 증거를 기반으로 하는 중재의 구축과 관련된 연구들이 활발하게 진행되어 왔고, 많은 연구자 및 교사들이 학생들의 학습 특성을 고려하여 무엇이 학생에게 효과적인 중재인지 파악하기 위해 노력하고 있다(Baron, 2004). 특히 증거기반실제의 구축을 위해 중재의 결과 및 효과크기 분석과 더불어, 각 중재가 체계적인 연구 설계 및 방법으로 수행되었고 그를 통해 효과를 입증하고 있는지를 살펴보는 것이 중요하다고 강조된다(서유진, 나경은, 2012).

연구와 그 연구에서 적용하고 있는 중재가 과학적 근거에 기초하고 있는지 확인하기 위해서는 먼저 개별 연구들의 연구설계 측면에서의 질을 평가할 필요가 있다(Odom et al., 2005). 이러한 질적 분석을 위하여 다양한 질적 지표(quality indicators)가 고안되어 활용되고 있는데, 대표적으로 미국 특수아동협회(Council for Exceptional Children, CEC)에서 발표한 특수교육 분야의 네 가지 연구 설계(실험집단 연구, 상관연구, 단일대상연구, 질적연구)별 질적 지표가 있다. CEC의 질적 지표 중 집단실

험설계(group experimental design) 및 준실험설계(quasi-experimental design) 연구에 대한 지표가 Gersten 외(2005)에 의해, 단일대상연구를 위한 지표가 Horner 외(2005)에 의해 개발되었다. (준)집단실험설계에 대한 Gersten 외(2005)의 질적 지표는 네 가지 영역으로 구성되는데, (1) 연구대상자, (2) 중재 실행 및 비교집단 기술, (3) 결과 측정도구, (4) 자료 분석이 그것이다. 단일대상설계에 대한 Horner 외(2005)의 질적지표는 (1) 연구대상자 및 연구환경, (2) 종속변수, (3) 독립변수, (4) 기초선, (5) 실험통제 및 내적타당도, (6) 외적타당도, (7) 사회적 타당도의 일곱 가지 영역으로 구성된다. 구체적인 지표는 <표 Ⅱ-4>에 제시하였다.

<표 Ⅱ-4> Gersten 외(2005)와 Horner 외(2005)의 질적 지표

연구설계	분류	구체적 지표
(준)집단실험설계 (Gersten et al., 2005)	연구대상자	- 연구대상자의 장애, 어려움에 대한 정보 - 집단 간 동질성 확보 - 중재제공자(교사)에 대한 정보 및 조건 간 중재제공자의 동등성 확보
	중재 실행 및 비교집단 기술	- 중재에 대한 기술 - 중재과정 충실도에 대한 기술 및 측정 - 비교집단의 교수적 특성 기술
	결과 측정도구	- 적절한 측정도구 및 일반화된 수행에 대한 다양한 측정도구 사용 - 측정 시기의 적절성
	자료 분석	- 연구문제에 적합한 분석기법 및 분석단위의 적절성 - 효과크기 산출
단일대상설계 (Horner	연구대상자 및 연구환경	- 연구대상자에 대한 정보 - 연구대상자 선별 과정 - 물리적 환경 특성에 대한 정보

et al., 2005)	종속변수	- 종속변수에 대한 기술 - 종속변수 측정 과정(측정가능한 지수 산출) - 측정의 타당성 및 기술 - 반복적 측정 - 측정신뢰도 또는 관찰자간 일치도 수집
	독립변수	- 독립변수에 대한 기술 - 독립변수의 체계적 조작 - 중재충실도 측정
	기초선	- 종속변수의 반복적 측정 및 안정화 - 기초선 기간에 대한 서술
	실험 통제 및 내적타당도	- 실험효과 입증 - 내적타당도에 대한 위험요소 통제 - 결과 패턴의 증명
	외적타당도	- 실험효과의 연구대상자 간, 환경 간, 자료 간 반복 가능성
	사회적 타당도	- 종속변수의 사회적 중요성 - 중재를 통한 종속변수의 변화의 사회적 중요성 - 독립변수 실행의 실용성 및 효과성 - 독립적인 수행에 대한 정보 제공(일반적 상황 속 일반적 중재자에 의한 향상 가능성)

2) 메타분석을 통한 효과크기 산출

메타분석은 비슷한 연구 주제에 대한 두 개 이상의 개별적인 연구결과들을 종합하여 하나의 효과크기로 통합하는 방법이다(Higgins et al., 2019). 이는 체계적 문헌고찰에서 가장 일반적으로 사용되는 양적 종합 방법으로서, 연구의 특성, 결과 등에 대한 체계적 요약(structured

summary)과 같은 서술적 종합과 구분된다(Higgins et al., 2019). 메타분석에서는 개별 연구에서 유의확률을 넘어 효과크기 산출 및 서로 다른 효과크기의 통계적 종합이 가능하다(Borenstein et al., 2011). 또한 개별 연구결과들의 동질성을 검증하고, 만일 이질성이 존재한다면 그 차이에 영향을 미치는 잠재적인 변인들을 확인할 수 있다(Higgins et al., 2019). 일반적으로 연구에 따라 연구대상자, 중재에 대한 조작적 정의, 측정도구 등의 특성이 서로 상이하므로, 개별적인 연구 결과만으로는 중재의 효과에 대한 정확한 답을 구하기 어렵다. 반면, 메타분석은 개별적인 연구결과에서는 도출할 수 없는 정보에 대한 답을 줄 수 있고, 보다 높은 통계적 정밀성과 검증력을 바탕으로 실제 효과를 확인할 수 있는 가능성을 높여준다(Higgins et al., 2019).

한편 메타분석의 장점에도 불구하고, 메타분석 기법에 의존하는 것에 대한 비판 또한 존재하는데, 예컨대 편향된 연구결과 수집으로 인한 잘못된 결론이 도출될 수 있다는 점, 독자적인 특성을 가진 개별 연구들을 하나로 종합하는 점 등이 있다(Higgins et al., 2019). 본 연구에서는 학령기 학습곤란 학생 대상 데이터기반교수의 효과를 종합하는 데 있어 이러한 제한점을 해소하기 위해 다음의 절차를 거쳤다. 메타분석을 실시하기에 앞서 분석대상논문들의 구체적인 특성을 살피고, 질적 지표를 활용하여 연구의 질을 검토하였다. 더불어 구체적인 포함배제준거에 따라 분석대상논문을 선정하였으며, 코딩 과정에서 발생할 수 있는 주관성을 배제하기 위하여 동료 연구자의 검토를 거쳤다.

5. 학습곤란 학생 대상 데이터기반교수의 선행연구 고찰

1) 미국의 선행연구 고찰

미국에서는 학습곤란 학생 대상 데이터기반교수를 주제로 몇 편의 문

헌 종합 및 메타분석 연구가 실시되었고, 그 중 대표적인 세 편을 개관하였다.

첫째, Jung 외(2018)에서는 여러 기초학습영역에 걸쳐 심각한 학습 요구(intensive learning needs)가 있는 학생들을 대상으로 증거기반개별화교수의 효과를 양적으로 종합하였다. 이를 위해 14편의 연구로 메타분석을 실시하였다. Jung 외(2018)의 연구에서는 분석대상논문 선정 시 상당히 엄격한 기준을 적용했는데, 예컨대 개별 연구에서 진전도 모니터링과 교수적 의사결정 두 가지 모두를 목적으로 CBM 검사를 활용해야 하고, 중재 및 모니터링을 7주 이상 동안 실시해야 하는 등 기준이 있었다. 연구 결과에 따르면 읽기, 쓰기, 수학 영역에 걸쳐 증거기반개별화교수의 효과를 확인한 결과 평균효과크기는 g = 0.37 ~ 0.38 (Hedge's g) 수준으로 나타났다. 더불어 예측변인의 조절 효과 분석 결과, 증거기반개별화교수가 학업 성취에 미치는 영향은 측정에 활용된 CBM 검사 종류(교사개발 CBM 검사를 사용했을 때 연구자 개발 경우보다 더 큰 효과크기가 나타남), CBM 측정 주기(한 주에 두 번씩 실시한 경우 가장 큰 효과크기가 나타남), 교사 지원의 유형과 빈도(교사에게 개별적인 컨설팅 및 소집단 지원을 함께 제공한 경우, 더욱 빈번하게 지원을 제공한 경우 큰 효과크기가 나타남)에 따라 상이한 것으로 나타났다. 여러 성취 결과들과 CBM 실시자에 따른 효과 차이는 유의하지 않았다.

둘째, Filderman 외(2018)의 연구에서는 증거기반 의사결정 절차를 포함한 읽기 중재 연구에 대한 문헌종합 및 메타분석을 실시하였다. 유·초·중등(K-12) 학년의 학습곤란 학생들을 대상으로 한 15개의 학술지 논문을 선정하여 분석한 결과, 작은 정도의 양적 효과크기를 확인할 수 있었다. 한편 Filderman 외(2018)에서 특히 주목할 만 한 점은 메타분석에 앞서 개별 연구가 미국 NCII(National Center on Intensive Intervention)의 증거기반개별화교수 모델에 부합하는 정도와 질적 지표를 충족하는 정도를 평가하였다는 것이다. 연구 결과에 따르면 교수적 수정과 관련하여 구조적인 조정(structural adjustments)에 비해 개별화된 조정을 할 때 증거기반 의사결정 읽기 중재의 효과가 높았다. 다만 분석대상논문

중 대부분에서는 데이터 수집 및 의사결정 과정에서 표준화된 CBM 검사도구가 아닌 숙달도 측정(mastery measurement)에 의존하였고, 한 편을 제외한 모든 연구에서 의사결정 주기 등 교육적 의사결정 과정에 대한 충분한 정보를 제시하지 않는 등 제한점이 있었다.

이 외에도 Stecker, Fuchs, & Fuchs(2005)는 장애학생의 성취를 증진하기 위한 CBM 활용의 효과에 대해 선행 연구를 고찰하고 효과를 검증하였다. 특히 체계적인 증거기반 의사결정 규칙 준수, 기술 분석(skills-analysis)을 바탕으로 한 성취에 대한 피드백 제공, 교수적 수정을 위한 지침 제공 등은 CBM 활용 효과를 잠재적으로 높이는 변인으로 볼 수 있다고 제시하였다.

2) 국내 선행연구 고찰

여러 문헌종합 및 메타분석이 이루어진 미국에 비해 국내에서는 학습곤란 학생 대상 데이터기반교수의 효과를 체계적으로 종합한 연구가 드문 편으로 나타났다.

먼저 여승수, 홍성두, 손지영(2014)의 연구에서 1999년부터 2014년까지의 국내 CBM 연구 동향을 검토한 바 있다. 연구에 따르면 27개의 학술지 논문 중 대다수는 2003년 이후에 출판되었으며, 19개에 달하는 대부분의 연구에서 기초학습기능 수행평가체제(BASA) 읽기 검사(김동일, 2000)와 수학 검사(김동일, 2006)를 CBM으로 사용하고 있었다. 이 외의 연구들은 특히 수학과 쓰기 영역에서 연구자가 직접 개발한 CBM 검사도구를 사용한 것으로 나타났는데, BASA 검사를 활용한 연구에 비해 신뢰도 및 타당도가 상대적으로 낮거나 관련 정보를 제시하지 않는 등 측정도구로서 제한점을 지니고 있었다. 또한 연구에서는 정보의 누락을 지적했는데, 전반적으로 20편에서만 신뢰도를, 16편에서만 타당도를 제시하는 등 상당수의 연구들에서 사용된 CBM 검사의 신뢰도와 타당도에 대한 충분한 정보를 제공하지 않았다.

이 외에도 정평강(2018)은 문헌 종합 연구를 통하여 장애로 진단 받거

나 위험군에 해당되는 학생을 대상으로 기초학습영역에서의 증거기반 개별화교수의 개념, 특성, 효과를 검토하였다. 다만 최종적으로는 문헌 선정 기준에 따라 국외 연구들만이 분석에 포함되었다. 이렇듯 국내 선행연구를 고찰한 결과, 학습곤란을 지닌 학령기 학생을 대상으로 기초학습영역의 데이터기반교수를 실시한 연구들을 보다 체계적으로 종합할 필요성이 있음을 확인할 수 있다.

한편 비록 데이터기반교수의 개념 자체에 초점을 두지는 않았지만, 국내에서도 현재까지 학습에 심각한 어려움을 겪는 학생들을 위해 집중적인 중재 프로그램을 적용한 뒤 효과를 확인하고 종합하려는 시도들은 꾸준히 있어왔다. 구체적으로 읽기(길한아 외, 2018; 김화수, 김민정, 박정숙, 2017; 김민경, 박유정, 2015; 김경미, 2013; 김우리, 고은영, 2012), 수학(전윤희, 2016; 김지원, 2008; 탁효정, 2017; 조은혜, 2017; 신미경, 채수정, 정평강, 2018; 김병룡, 2017), 쓰기(정광조, 홍성두, 2017; 고서연 외, 2018; 김동일, 이재호, 이미지, 2013) 영역 각각에서 중재 효과를 종합하는 메타분석 연구가 실시되었다. 그러나 이러한 선행연구들은 대부분 특정 기초학습영역에 국한하거나, 구체적인 중재 프로그램 또는 전략의 효과크기를 확인하였고, 중재 프로그램 또는 전략의 선택 이후 그 적용 과정에서 데이터기반교수의 효과를 검증하는 데 목적을 두지는 않았다.

선행연구 고찰을 통해 데이터기반교수가 효과적인 접근으로 꾸준히 제시되어 온 것에 비해 국내 연구에서 이에 초점을 두고 그 경험적 증거를 제시하고자 한 연구가 부족한 실정임을 알 수 있다. 즉 국내에서 학습곤란을 지닌 학령기 학생을 대상으로 실시된 데이터기반교수 연구를 체계적으로 종합할 필요성이 있다. 이를 통해 국내의 교육 현장에서 여러 가지 이유로 심각한 학습 문제를 경험하는 학령기 학생들을 위해 기초학습기술을 가르칠 때 데이터기반교수를 사용하는 것이 타당한지 확인하고, 그 효과를 체계적으로 종합하여 제시할 필요가 있다.

3) 국내 (준)집단실험설계 연구의 질적 수준 및 효과크기 분석

2020년 12월까지 국내에서 집단실험설계 또는 준집단설계의 형태로 학령기 학습곤란 학생을 대상으로 데이터기반교수를 실시한 연구를 살펴보면 다음과 같다. 기초학습영역에서 데이터기반교수를 적용하고, 그 중에서도 표준화된 CBM 도구인 BASA를 활용하여 반복 측정 및 진전도 모니터링을 한 연구를 선정하였다. '데이터기반교수'의 용어를 명시하지 않더라도 주요 절차를 따르고 있다고 판단될 경우 분석에 포함하였다. 구체적인 검색어 및 문헌 선정 기준은 단일대상설계 연구의 경우와 동일하고, 이는 이후에 제시되는 연구방법에서 보다 자세히 제시하였다. 이러한 절차에 따라 수집된 (준)집단실험설계 연구는 총 6편이었다.

먼저 연구의 주요 특징은 <표 Ⅱ-5>과 같다. 수학 영역에서 다섯 편, 쓰기 영역에서 한 편의 연구가 이뤄졌고, 모두 일반학교의 초등학생을 대상으로 하고 있었다. 6편의 연구 중 세 편의 연구(이새별 외, 2015; 김혜영, 강옥려, 2010; 홍성두, 여승수, 2011)에서는 1단계 일반교수-2단계 소집단 교수-3단계 개별교수로 이어지는 중재반응모형(RTI)에 기초한 중재를 제공하고 있었는데, 이 경우 소집단교수와 개별교수에 해당되는 2단계 또는 3단계 중재를 기준으로 데이터를 코딩하였고, 사전점수와 사후점수가 보고되지 않은 경우 중재가 제공된 첫 주의 점수와 마지막 주의 점수를 코딩하였다.

<표 Ⅱ-6>에서 확인할 수 있듯이 데이터기반교수와 관련된 특징은 연구마다 상이했는데, 예컨대 CBM 측정도구를 통한 진전도 측정 주기는 매 회기부터 매 8회기까지 광범위했다. 중재반응모형을 적용한 세 편의 연구에서는 나머지 연구에 비해 데이터기반교수 절차와 관련된 정보를 비교적 충실하게 제시하고 있었다. 구체적으로는 다음 단계의 집중중재가 필요한지를 판단하기 위해 학생의 데이터를 활용하였고, 비록 선행연구에서 보고되는 데이터기반 의사결정 유형에는 해당되지 않지만 다음 중재 단계를 결정하기 위해 수행수준 및 진전도가 모두 낮음을 확인하는 이중불일치 준거에 기초하고 있었다. 다만 수행수준의 불일치는 주로 백분위 점수에 기초한 것으로 밝히고 있었으나, 진전도의 경우 성장선의 기울기, 데이터포인트 개수 등 구체적으로 어떤 방법으로 불일치를 확인

하였으며, 어느 정도의 진전이 보이지 않을 경우 의사결정을 내렸는지 등은 기술하지 않아 제한점이 있었다. 또한 해당 연구에서는 학생의 데이터에 기초하여 중재 단계를 변화시킴에 따라 중재 횟수, 회기 당 수업 시간, 교수집단 크기, 수업 형태, 새로운 교수 전략 등 양적, 질적으로 교수적 수정을 가하였음을 확인할 수 있었다. 다만 나머지 세 편의 연구에서는 데이터기반 의사결정 준거 및 교수적 수정에 대한 정보를 충분히 제시하지 않은 것으로 분석되었다.

다음으로 (준)집단실험설계 연구를 중심으로 학령기 학습곤란 학생 대상 데이터기반교수의 효과크기를 살펴보면 다음과 같다. 6편의 연구결과에서 산출된 효과크기를 종합하기 보다는 개별적인 효과크기를 산출하여 제시하였고, 교정 효과크기인 Hedge's g를 사용하였다. 이는 표본 집단의 수가 비교적 작거나 각 연구에 포함된 집단의 크기가 서로 다를 때 Cohen's d에 비해 정확한 효과크기를 도출하는 것으로 알려져 있다(Hedges, 1981; Hedges & Olkin, 1985). 연구설계에 따른 Hedge's g 산출 공식은 다음과 같다.

첫째, 실험집단(T)과 통제집단(C)이 각각 있는 실험통제집단 사후검사 설계의 경우, Hedge's g는 아래의 공식으로 Cohen's d로부터 산출된다(Hunter & Schmidt, 2004; Borenstein et al., 2011). $\overline{X}_T$와 $\overline{X}_C$는 각각 실험집단과 통제집단의 사후검사 평균 점수를 의미하고, SD_{pooled}는 두 집단 사후검사 점수의 통합 표준편차(pooled standard deviation)를 의미한다. N_T과 N_C는 실험, 통제집단 각각의 표본 크기를 의미하며, SD_T^2와 SD_C^2는 각 집단에서 사후검사 점수의 분산을 나타낸다. 이를 통해 산출된 d의 분산 또한 다음과 같이 계산할 수 있다.

$$d = \frac{\overline{X}_T - \overline{X}_C}{SD_{pooled}} \quad (d = \text{Cohen's } d)$$

$$SD_{pooled} = \sqrt{\frac{(N_T - 1)SD_T^2 + (N_C - 1)SD_C^2}{N_T + N_C - 2}}$$

$$V_d = \frac{1}{N_T} + \frac{1}{N_C} + \frac{d^2}{2(N_T + N_C)}$$

다음으로 Cohen's d로 계산된 표준화된 평균 효과크기를 Hedge's g로 변환하는 방법은 다음과 같다. 여기서 교정지수 J는 Cohen's d가 효과크기를 과대 추정하는 경향을 교정하는 역할을 한다. Hedge's g의 분산역시 아래의 계산식을 통해 계산할 수 있다.

$$g = J \times d \quad (g = \text{Hedge's } g)$$

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

$$V_g = J^2 \times V_d \ , \ SE_g = \sqrt{V_g}$$

둘째, 실험통제집단 사전-사후검사 설계에서는 실험통제집단 사후검사 설계의 경우와 동일한 계산식이 사용되나, 사전-사후검사의 평균 차이 점수를 활용한다는 점에서 차이가 있다(Morris, 2008). 아래의 계산식에서 $\overline{X}_{T,pre}$, $\overline{X}_{T,post}$, $\overline{X}_{C,post}$, $\overline{X}_{C,post}$ 는 각각 실험집단, 통제집단의 사전점수와 사후점수를 나타내고, 사전점수의 통합 표준편차를 나타내는 SD_{pooled} 또한 아래 공식으로 계산된다(Hunter & Schmidt, 2004; Borenstein et al., 2011). 이후에는 동일한 방식으로 Cohen's d에서부터 Hedge's g를 계산한다.

$$d = \frac{(\overline{X}_{T,post} - \overline{X}_{T,pre}) - (\overline{X}_{C,post} - \overline{X}_{C,pre})}{SD_{pooled}} \quad (d = Cohen's d)$$

$$SD_{pooled,pre} = \sqrt{\frac{(N_T - 1)SD^2_{T,pre} + (N_C - 1)SD^2_{C,pre}}{N_T + N_C - 2}}$$

$$V_d = \frac{1}{N_T} + \frac{1}{N_C} + \frac{d^2}{2(N_T + N_C)}$$

셋째, 통제집단이 없는 단일집단 사전사후검사의 경우 다소 다른 계산 공식이 활용되는데, 이때에는 통합 표준편차 계산 시 사전점수와 사후점수 간 상관계수(r)가 고려된다. 상관계수의 값은 여러 가지로 설정할 수 있는데, 일반적으로 0.5로 설정한다. d 값과 통합 표준편차, d의 분산은 아래와 같다. 나머지 계산식은 앞선 두 집단 실험설계와 동일하다. Hedge's g의 크기는 Cohen의 효과크기 해석기준에 따라 작은 효과크기(0.2), 중간 효과크기(0.5), 큰 효과크기(0.8)로 해석할 수 있다(Cohen, 2013). 여섯 편의 연구별 효과크기는 통계프로그램인 CMA(Comprehensive Meta-Analysis)를 사용하여 분석하였다.

$$d = \frac{\overline{X}_{post} - \overline{X}_{pre}}{SD_{pooled}} \quad (d = \text{Cohen's } d)$$

$$SD_{pooled} = \frac{SD_{diff}}{\sqrt{2(1-r)}}$$

$$S_{diff} = \sqrt{S_{pre}^2 + S_{post}^2 - (2 \times r \times S_{pre} \times S_{post})}$$

$$V_d = (\frac{1}{N} + \frac{d^2}{2N}) \times 2(1-r)$$

분석 결과, 6편의 (준)집단실험설계에서 도출된 효과크기의 범위는 0.752 ~ 3.256로 매우 넓었고, 평균값은 1.54, 중앙값은 1.20으로 나타났다. Cohen(2013)의 해석기준에 따라 0.8 이상은 큰 효과크기에 해당되는데, 이러한 결과는 매우 높은 효과크기로 해석된다. 분석대상논문 중 김동일, 이태수(2005)의 경우, 두 개의 실험집단 중 직접교수와 진전도 모니터링을 함께 실시한 실험집단을 기준으로 효과크기를 계산하였는데, 초등학교 2학년과 3학년을 대상으로 각각 3.256, 2.979의 매우 높은 효과크기가 산출되었다. 해당 연구에서는 충분한 경험적 증거가 확립된 교수전략인 직접교수가 적용된 교재의 활용, 주기적인 진전도 측정, 오류에 대한 피드백 등 데이터기반교수의 원리를 보다 체계적으로 따르고 있다는 데 의의가 있지만, 추후 집단설계 연구의 효과크기를 종합할 때 이상치로 고려될 필요가 있다.

<표 Ⅱ-5> 집단설계 연구의 주요 특징

저자 (연도)	출판 유형	연구 설계	학생유형		표본크기 (남, 여)	학교급/ 학교유형	중재내용(독립변인)	비고
			실험	비교				
이태수 (2010)	학술지	실험통제 사전사후	쓰기학습 장애	쓰기학습 장애	40 (25,15)	초 4,5 /통합학급	아동문학작품을 활용한 감상문 표현활동	비교집단: 전통적 교수
김동일, 이태수 (2005)	학술지	실험통제 사전사후	수학학습 부진/학습 장애	수학학습부진/학습장애	80 (미제시)	초 2,3 /통합학급	직접교수 및 CBM에 근거한 수학 BASA 모니터링	비교집단: 전통적 교수
강옥려, 박보영 (2010)	학술지	단일집단 사전사후	수학학습 장애위험	-	12 (7,5)	초 1 /통합학급	활동중심 수감각 학습	
이새별 외(2015)	학술지	다단계 성장모형	수학학습 부진(2단계)/학습 장애(3단계)	-	25명(2단계), 7명(3단계)	초 3 /통합학급	중재반응모형에 의한 3단계 직접교수 중재프로그램 (3단계 Touchmath 전략 활용)	1단계 일반교수, 2단계 소집단교수, 3단계 개별교수
김혜영, 강옥려 (2010)	학술지	다단계 성장모형	수학학습 부진(2단계)	-	13명(5,8)	초 2 /통합학급	중재반응모형에 의한 2단계 수학 중재 (구체물-반구체물-추상물 활용)	1단계 일반교수, 2단계 소집단교수
홍성두, 여승수 (2011)	학술지	잠재성장 계층모형	연산학습 부진(2단계)	-	25 (미제시)	초 3 /통합학급	중재반응모형에 의한 2단계 수학 중재	

<표 Ⅱ-6> 집단설계 연구의 데이터기반교수 관련 특징 및 효과크기

저자 (연도)	중재회기수	회기당 시간(분)	중재자	진전도 측정주기	증거기반 의사결정	종속변인	Hedge's *g*	SE	비고
이태수 (2010)	16 (8주, 주2회)	40	또래	8회기	미제시	쓰기	1.195	0.338	정확음절수
							1.559	0.356	어휘수
							1.004	0.330	문장수
							0.901	0.326	문장길이
김동일, 이태수 (2005)	12 (6주, 주 2회)	35~45	학습부진반 교사	2~3회	미제시	기초연산	3.256	0.551	2학년
							2.979	0.572	3학년
강옥려, 박보영 (2010)	10 (4주, 주 3회)	40~50	연구자	2회기	미제시	수학성취도	1.849	0.463	
이새별 외(2015)	10회기 (4주, 주 3회)	40	학습부진반 교사	매회기	(이중불일치)	연산능력	0.752	0.221	2단계
	12회기 (4주, 주3회)	40	연구자	1~2회	(이중불일치)	연산능력	0.974	0.419	3단계
김혜영, 강옥려 (2010)	36회기 (9주, 주4회)	40	연구자, 담임교사	4회기	(이중불일치)	연산능력	1.528	0.397	2단계
홍성두, 여승수 (2011)	10회기 (4주, 주2~3회)	40	학습부진반 교사	매회기	(이중불일치)	연산능력	0.920	0.233	2단계

다음으로 총 6편의 연구를 Gersten 외(2005)가 제시한 질적 지표에 기초하여 분석한 결과는 <표 Ⅱ-7>과 같다. 분석은 3점 척도로 이루어졌고, 2점 이상의 평균을 나타냈을 때 충분한 질적 수준을 갖춘 것으로 해석하였다.

첫째, 연구대상자 영역에서 6편의 연구 모두 연구대상자의 정보를 충분하게 제시하고 있는 것으로 평가되었다. 연구대상자의 집단 간 동질성 지표의 경우 비교집단을 따로 설정하지 않은 4편의 연구에 대해서는 중간 점수로 평가하였다. 나머지 연구들에서 모두 각 집단에 포함된 학생들의 지능지수, 표준화된 성취검사를 통한 현재 성취 수준 등 정보를 제공함으로써 집단 간 동질성을 확보하고 있는 것으로 평가되었다. 중재제공자의 정보 부분에서는 충분하게 충족되는 경우가 1편, 부분적으로 충족되는 연구가 3편, 충족되지 않는 연구가 1편으로 평가되었다. 특히 대부분의 연구에서 비교집단과 실험집단 교수에 참여한 중재제공자의 정보를 비교하여 제시하지 않고 있었고, 중재제공자에 대한 기본적인 정보를 밝히지 않은 경우도 있어 보다 나은 질적 수준의 확보가 필요한 것으로 나타났다.

둘째, 중재 실행 및 비교집단 영역에서 6편의 연구 모두 중재에 대해 구체적이고 명확하게 설명하고 있었으나, 중재충실도를 측정한 연구가 2편에 불과하고 나머지 연구에서는 이를 제시하지 않고 있었다. 비교집단에 제공된 교수 관련 기술의 경우, 단일집단 대상으로 중재를 제공한 연구의 경우 중간 점수로 평가하였고, 그 외 나머지 연구에서는 충족하고 있는 것으로 나타났다.

셋째, 결과 측정도구와 관련하여, 6편의 연구에서 모두 적절한 측정도구를 사용하고 있었으나, 두 개 이상의 측정도구를 활용하여 일반화된 수행에 대한 정보를 제공한 연구는 1편에 불과하였다. 자료 수집 시기의 적절성에 대해서는 6편에서 모두 충족하는 것으로 나타났다.

넷째, 자료 분석 영역에서는 6편의 연구 모두 연구문제에 적절한 분석기법 및 분석 단위를 사용하고 있었다. 다만 모든 연구에서 효과크기를 계산하지 않고 추론통계의 유의확률을 제시하거나 대안적인 방법으로 결

과의 변화 크기를 제시하는 데 그치는 것으로 평가되었다.

<표 Ⅱ-7> 집단설계 연구의 질적 지표 분석 결과

분류	구체적 지표	논문 수			평균
		1 (미충족)	2 (부분충족)	3 (충족)	
연구대상자	연구대상자의 장애, 어려움에 대한 정보를 충분히 제공하였는가?	0	0	6	3.00
	집단 간 연구대상자의 동질성이 확보되었는가?	0	4	2	2.33
	중재제공자(교사)에 대한 정보를 충분히 제공하고, 조건 간 중재제공자의 동등성을 확보하였는가?	2	3	1	1.83
중재 실행 및 비교집단 기술	중재가 명백하고 자세하게 기술되었는가?	0	0	6	3.00
	중재과정 충실도가 기술되고 측정되었는가?	4	0	2	1.67
	비교집단에 제공된 교수, 서비스 등 특성을 기술하였는가?	0	4	2	2.33
결과 측정도구	중재에 적절한 측정도구를 사용하고, 일반화 수행의 도구 간 균형을 위해 여러 측정도구를 사용하고 있는가?	0	5	1	2.17
	결과는 적절한 시기에 측정되었는가?	0	0	6	3.00
자료 분석	연구문제에 적합한 분석기법이 사용되고, 분석단위가 적절히 설정되었는가?	0	0	6	3.00
	추리통계 뿐 아니라 효과크기 계산을 포함하였는가?	0	6	0	2.00

Ⅲ. 연구방법

본 연구에서는 단일대상설계 연구를 중심으로 국내 학령기 학습곤란 학생 대상 데이터기반교수 연구의 질과 중재의 효과를 확인하고자 하였다. 이러한 목적 하에 다음의 방법이 활용되었다. 첫째, 데이터기반교수 연구의 연구설계 및 방법 측면의 질적 수준을 확인하기 위하여 Horner 외(2005)가 제시한 연구설계에 따른 질적 지표 및 관련 선행연구들에 기초하여 질적 분석을 실시하였다. 둘째, 학령기 학습곤란 학생에게 데이터기반교수가 미치는 효과를 양적으로 종합하기 위하여 2수준 다층메타분석을 실시하였다. 구체적으로 전체 효과크기를 도출하고, 연구참여자 간 효과크기의 이질성을 검정하였으며, 연구참여자 수준에서 효과크기의 이질성을 설명하는 변인을 탐색하였다. 이와 같은 절차는 다음의 자료 수집, 포함 및 배제준거 설정, 데이터 코딩, 데이터 분석으로 진행되었다.

1. 자료 수집

먼저 국내 주요 논문검색사이트인 학술연구정보서비스(RISS), 한국학술정보(KISS), 누리미디어(DBPIA) 등을 활용하여 2020년 12월까지 출판된 문헌을 체계적으로 수집하였다. 학습곤란 학생을 위한 주제어(‘학습곤란’, ‘학습장애’, ‘학습장애 위험’, ‘저성취’, ‘학습부진’ 등), 기초학습능력을 위한 주제어(‘기초학업’, ‘기초학습’, ‘읽기’, ‘쓰기’, ‘수학’ 등), 데이터기반교수를 위한 주제어(‘데이터기반교수’, ‘증거기반교수’, ‘데이터기반 중재’, ‘데이터기반 개별화’, ‘교육과정중심측정’, ‘CBM’ 등), 그리고 BASA를 위한 주제어(‘기초학습기능 수행평가체제’, ‘BASA’ 등)을 검색어로 사용하여 여러 조합으로 검색하였다. 또한 누락을 최소화하기 위하여 기존의 문헌종합연구와 분석대상논문들의 참고문헌을 검토하며 추가하는 과정을 거쳤다. 한편 본 연구에서는 개별 연구에서 ‘데이터기반교수’의 용어 및 개념을 명시하고 있지는 않지만 BASA를 활용하여 데이터기반교수를 충

분히 따르고 있다고 판단될 경우 분석에 포함하였다. 이는 BASA가 CBM 원리에 기초하여 개발되어, 그 활용 방식에 이미 진전도 모니터링, 학생 데이터를 활용한 교수적 의사결정 및 목표 조정 등 데이터기반교수의 개념 및 절차를 내포하고 있기 때문이다.

이러한 자료 수집 절차를 통해 중복을 제외한 후 총 422편의 연구가 일차적으로 선정되었다. 이 중에서 390편의 연구는 다음의 포함 및 배제 준거에 의해 최종 분석대상에서 제외되었다.

2. 포함 및 배제준거

최종 분석대상논문을 선정하기 위하여 본 연구에서 활용한 포함 및 배제준거는 다음과 같다.

첫째, 학습곤란 학생을 연구참여자로 선정하고 있는 연구만이 포함되었다. 개별적인 연구에서 '학습장애', '학습장애 위험', '저성취', '학습부진', '학습곤란' 등으로 표현한 학생들은 모두 넓은 범주의 학습곤란 학생에 포함된다고 판단하였다. 본 연구에서는 지적장애를 학생을 대상으로 한 연구도 분석대상으로 포함하였다. 일반적으로 지적장애를 지닌 학생의 경우 학습곤란 범주에 포함되지 않는 것으로 본다(Elkins, 2000). 그러나 이들의 특수한 교육적 요구 또한 학습곤란의 관점에서 이해할 수 있으며, 특수교육대상자의 약 70%가 일반학교에 배치되어 있고 일반학교에서 학습 상 문제가 있는 학생들 중에는 경도지적장애 및 경계선급 지능 등 특성을 가진 학생들이 포함되어 있다고 보고되므로(이대식, 2019), 본 연구에서는 이들을 대상으로 한 데이터기반교수의 효과도 확인하고자 하였다.

둘째, 6세부터 18세의 학령기 학생을 대상으로 한 연구를 포함하였고, 해당 연령 범위를 벗어난 학생을 대상으로 하는 경우 분석에서 제외하였다.

셋째, 기초학습영역(읽기, 쓰기, 수학)에서 중재를 실시하고 종속변수

에 기초학습능력과 관련된 변수를 포함하여 측정하고 있는 논문을 선정하였다. 독립변수로서 구체적인 중재의 유형 및 내용에는 제한을 두지 않았다. 한편 연구에 따라 특정한 중재를 제공한 후 그 효과를 검증하는 데 목적이 있는 것이 아니라 잠재계층분석을 통한 집단 특성 확인, 중재반응모형(RTI)의 적용 가능성 탐색, 학습장애아동 선별 등 다른 목적 하에 수행된 경우가 있었는데, 이 경우 구체적인 내용을 확인한 후 포함 여부를 결정하였다. 구체적으로 기초학습영역에서 중재가 실시되었고, CBM 측정도구인 BASA를 활용하여 진전도 모니터링 등 데이터기반교수에 적합한 절차를 거쳤으며, 연구참여자 및 종속변수별 효과크기 산출을 위한 충분한 통계적 수치가 제시되어 있을 경우 분석에서 제외하지 않았다.

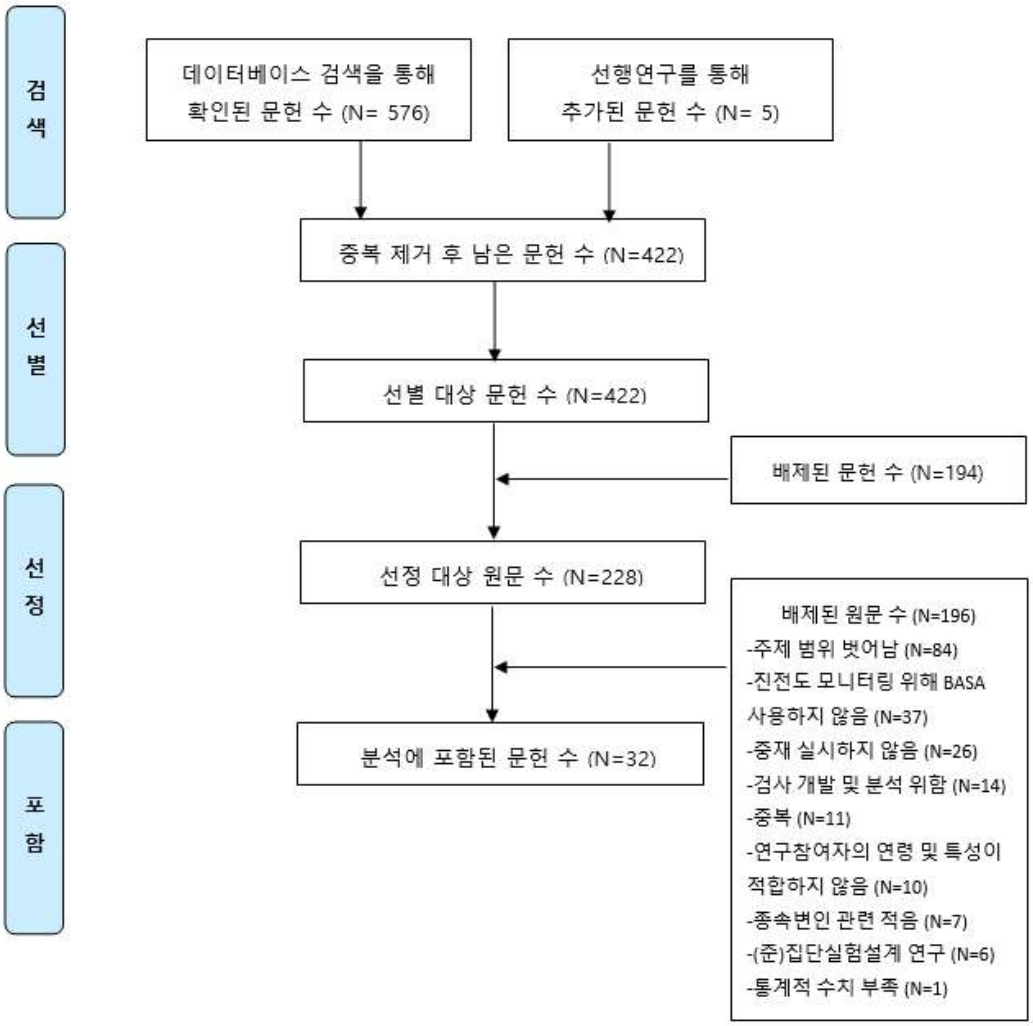
넷째, 본 연구의 목적은 데이터기반교수의 효과를 확인하는 데 있으므로, 데이터기반교수의 개념 및 절차를 적용한 연구만을 포함하였다. 선행연구는 효과적이고 충실한 데이터기반교수가 포함해야 하는 절차들을 구체적으로 제시하고 있다. 그러나 현재까지 국내에서 데이터기반교수 자체에 집중하여 그 절차를 명시하며 중재를 실시한 연구가 많지 않은 실정이었다. 또한 그 개념을 명시하지 않았더라도 CBM을 활용한 주기적인 진전도 모니터링, 학생 데이터에 의한 지속적인 교수적 수정 및 의사결정 등 데이터기반교수가 적용되었다고 판단할 수 있는 연구들이 상당수 있었다. 이에 본 연구에서는 첫째, BASA를 활용한 반복 측정, 둘째, 교수적 수정 및 교육적 의사결정을 위한 학생의 진전도 모니터링의 두 가지를 데이터기반교수의 필수적인 절차로 판단하고, 이를 적용한 경우 분석에 포함하였다.

다섯째, CBM 도구로서 BASA를 사용한 논문을 분석대상에 포함하였는데, 이는 BASA가 국내 기초학습영역에서 측정학적 적합성이 검증되고 상용화된 표준화 CBM 측정도구이기 때문이다. Jung 외(2018)에 따르면 연구자 또는 교사 개발 CBM을 사용하였을 때 검사 자체가 지니는 측정학적 적합성에 대한 근거가 부족하고, 진전도 모니터링 목적으로서 적합성이 검증되었다고 보기 어렵다. 특수교육 연구에서 신뢰도 및 타당

도는 매우 중요하며, 종속변수 측정에 있어 측정학적 적합성에 대한 증거가 확보되어 있는 측정도구를 사용하는 것은 (준)집단실험연구 연구설계의 질을 검토하는 지표 중 하나에 해당될만큼 강조된다(Gersten et al., 2005). 그러나 실제 CBM을 활용한 국내 연구들을 종합한 선행연구에 따르면 연구자 개발 도구를 사용한 많은 수의 연구에서 이러한 정보를 제공하지 않아 한계가 지적된 바 있다(여승수, 홍성두, 손지영, 2014). 따라서 연구자 또는 교사가 개발한 CBM 측정도구를 활용하여 진전도를 측정한 연구는 분석에서 제외하였고, 검사 실시 및 해석 절차가 표준화되어 있고 측정학적 적합성에 대한 충분한 근거가 확보된 BASA를 적용한 연구로 제한함으로써 데이터기반교수의 효과를 더욱 정확하게 분석하고자 하였다.

여섯째, 효과크기 계산을 위한 충분한 통계치를 제공하고 있는 단일대상설계 연구를 포함하였다. 단일대상설계 연구는 개별적인 연구 참여자를 중심으로 중재 효과를 검증할 수 있는 연구방법(이소현, 박은혜, 김영태, 2000)이다. 많은 수의 표본을 구하기 어렵거나 개별 대상의 특성을 고려하여 집중적인 맞춤형 중재가 필요한 경우 적절하며(Shadish, Rindskopf, & Hedges, 2008), 체계적인 실험 분석을 통해 연구대상 및 상황에 적합한 교수법을 찾아내고, 과학적 연구를 바탕으로 특수교육 현장의 실제를 확인, 개선하는 데 중요한 역할을 해왔다(한성희, 남윤석, 2005). 특히 연구참여자의 특수한 요구에 대한 개별화된 접근을 취하고 반복측정을 하는 데이터기반교수의 연구에 적절하다. 본 연구에서는 구체적인 연구 설계를 밝히고 있지는 않으나, 기초선, 중재 구간으로 판단할 수 있는 구분된 구간을 설정하고 있거나 반복 측정한 진전도 데이터를 시각적 그래프 등으로 나타내고 있는 단일사례연구도 포함하였다.

포함 및 배제준거에 따라 최종적으로 분석대상논문으로 선정된 단일대상설계 연구는 총 32편이었다. 이러한 분석대상논문 선정절차를 PRISMA 4단계 도식에 따라 제시하면 [그림 Ⅲ-1]과 같다.



[그림 Ⅲ-1] PRISMA 4단계 도식

3. 데이터 코딩

1) 질적 지표 분석을 위한 데이터 코딩

본 연구에서는 분석대상논문으로 선정된 연구들이 증거기반실제로서 수용 가능한 정도를 확인하고자 Horner 외(2005)가 제시한 단일대상설계

연구 질적 지표에 기초하여 질적 분석을 하였다. 다만 해당 질적 지표는 매우 엄격한 기준을 가지고 있고, 각 지표의 충족 수준을 단순히 이분법적으로 평가하기 어렵다는 제한점이 지적되어 왔다(안예지, 이미지, 2020; Jitendra, Burgess, & Gajria, 2011). 이에 본 연구에서는 Horner 외(2005)의 질적 지표에 기초하되, 질적 지표를 적용하여 증거기반실제 수준을 분석한 선행연구(이예다나 외, 2015; 안예지, 이미지, 2020; 최희승, 2020)를 참고하여 3점 평가척도 기준에 따라 분석하였다. 지표 및 영역에서 2점 이상의 평균을 보였을 때 충분한 질적 수준을 갖춘 것으로 해석하였다.

질적 지표 분석의 신뢰도를 높이기 위해 연구자 외에 Gersten 외(2005), Horner 외(2005)의 질적 지표에 대한 이해를 가지고 있는 특수교육전공 박사과정생 1인이 질적 분석에 참여하였다. 질적 지표에 대한 동일한 이해를 점검한 뒤 분석대상논문 중 집단설계연구 두 편과 단일대상설계 연구 두 편을 무작위로 선정하여 함께 분석하는 과정을 통해 각 지표의 내용 및 척도기준에 대한 이해가 일치되도록 하였다. 이후 박사과정생 1인은 연구자가 분석한 전체 질적 지표 분석 자료를 검토하였고, 의견이 불일치하는 항목을 확인하였다. 일치되지 않은 항목에 대해서는 추가적으로 검토하고 함께 충분한 논의를 통해 합의를 도출하였다. 분석자 간 의견이 일치된 평가치를 일치된 평가치와 일치되지 않은 평가치를 합한 수치로 나눈 뒤 100을 곱해 계산한 분석자 간 신뢰도는 99.6%로 나타났다. 본 연구에서는 질적 지표 분석을 위해 Chard 외(2009)의 루브릭을 참고하였고, 구체적으로 활용된 평가 루브릭은 <부록 1>과 <부록 2>에 제시하였다.

2) 메타분석을 위한 데이터 코딩

메타분석을 위한 코딩틀을 개발하기 위하여 데이터기반교수에 대한 메타분석을 실시한 선행연구(Filderman et al., 2018; Jung et al., 2018)의 코딩틀과 데이터기반교수 개념 및 절차의 주요 구성요소를 제시한 선행

연구(NCII, 2013; Stecker, Fuchs, & Fuchs, 2005; 정평강, 신재현, McMaster, 2016; 정평강, 2018)를 참고하였다. 본 연구에서 활용된 코딩틀은 <표 Ⅲ-1>과 같다.

더불어 분석자 간 신뢰도를 확보하기 위하여 마찬가지로 메타분석을 활용한 연구 경험이 있는 특수교육전공 박사과정생 1인이 메타분석을 위한 코딩에 참여하였다. 연구자는 코딩자에게 코딩틀의 내용과 코딩 방법을 설명한 후, 무작위로 선정한 세 편의 연구를 함께 코딩하였다. 이후 연구자 외 코딩자는 연구자가 코딩한 분석 자료를 검토하며 의견이 불일치하는 항목을 기록하였다. 데이터기반 의사결정 규칙, 교수적 수정 등 유형 분류에 있어 주관성이 개입될 수 있는 영역은 더욱 신중하게 살펴보았다. 이후 두 명의 코딩자는 불일치 항목에 대해 추가적인 검토와 논의를 통해 합의를 도출하는 과정을 거쳤다. 메타분석을 위한 데이터 코딩의 분석자 간 신뢰도는 98.86%로 나타났다.

<표 Ⅲ-1> 메타분석을 위한 코딩틀

구분	내용
일반적인 특징	
출판연도	연구가 출판된 연도 (2005~2020)
출판유형	(1) 학술지 (2) 학위
연구참여자 관련 특징	
학교급	연구참여자가 재학 중인 학교급 (1) 초등학교(저학년, 고학년) (2) 중학교 (3) 고등학교
표본 크기	연구참여자의 수
성별	연구참여자 중 남자, 여자의 수
학생유형	연구참여자의 곤란의 유형 (1) 학습장애 진단/위험 (2) 저성취 (3) 학습부진 (4) 지적장애 (5) 기타
학교유형	연구참여자가 재학 중인 학교 유형

	(1) 일반학급 (2) 특수학급 (3) 특수학교 (4) 혼합 (5) 기타
중재 관련 특징	
중재내용	기초학습영역 내 중재프로그램(독립변인)의 주요 내용 (예: 놀이활동을 이용한 수학학습)
교수전략	사용된 교수 전략의 유형 (예: 직접교수)
중재장소	중재가 제공된 장소 (1) 일반학급 (2) 특수학급 (3) 특수학교 (4) 치료실(사설센터) (5) 기타 (6) 미제시
중재 회기 수	제공된 총 중재 회기 수
회기 당 시간	한 회기 당 시간(분)
집단 크기	교수집단의 크기 (1) 일대일 (2) 소집단(2-3명) (3) 중집단(4-6명) (4) 미제시
중재 실시자	중재를 제공한 사람
기초학습영역	목표 기술 및 능력(종속변인)으로 설정된 기초학습영역 (1) 읽기 (2) 쓰기 (3) 수학 (4) 기타
증거기반 교수 관련 특징	
CBM 유형	CBM 도구로 사용된 BASA 검사의 유형 (읽기유창성, 초기문해, 쓰기, 초기수학, 수학 등)
진전도 측정 주기	CBM 도구를 활용한 진전도 측정 주기 (1) 매회기 (2) 2-3회기마다 (3) 4회기마다 (4) 기타
진전도 측정 실시자	진전도를 측정한 사람 (1) 중재자 (2) 중재자 이외 (3) 미제시
데이터기반 의사결정 규칙	학생의 진전도 데이터를 바탕으로 교수적 의사결정을 하기 위해 사용한 규칙

	(1) 기울기 방법 (2) 데이터포인트 방법 (3) 숙달기준 방법 (4) 기타 (5) 미제시
교수적 수정 유형	실시한 교수적 수정의 유형 (1) 양적 (2) 질적 (3) 양적+질적 (4) 미제시

4. 데이터 분석

기존에 메타분석은 주로 집단실험설계를 중심으로 이루어진 경향이 있는데, 최근 단일대상설계 연구를 중심으로도 개별적인 연구들의 결과를 종합하려는 필요성이 제기되었고, 이에 메타분석이 점차 활발하게 활용되어 왔다(Kratochwill & Levin, 2014). 그러나 아직까지 단일대상설계 연구 결과의 양적 종합에 있어 가장 적절한 통계분석방법이 무엇인지는 합의가 이뤄지지 않았다. 효과크기 계산에 있어서도 가장 정확한 효과크기 지수가 무엇인지 합의에 이르지 못하였으나(Gage & Lewis, 2014), 비중복비율(percentage of nonoverlapping data, PND), 개선율 차이(improvement rate difference, IRD), Tau-U 등이 단일대상설계 연구의 효과크기 지수로 빈번하게 사용되어 왔다.

한편 단일대상설계 연구에서 이러한 중재효과의 계산은 측정값의 분산 정도, 경향성, 중복비율 등의 시각적 분석에 의존해온 경향이 있다(Brossart et al., 2006; Van den Noortgate & Onghena, 2003a). 그러나 이는 다소 주관적이고 여러 단일대상설계 연구 간의 효과 비교가 어렵다는 제한점이 지적되었다(Wang, Cui, & Parrila, 2011). 이러한 측면에서 다층모형을 통한 분석은 단일대상설계 연구의 메타분석에 있어 또 하나의 주목받는 분석방법이다(Jenson et al., 2007; Van den Noorgate & Onghena, 2007). Van den Noortgate & Onghena (2003a, 2003b; 2007)는 여러 단일대상연구에서 중재효과를 양적으로 나타내고 종합, 비교하기 위한 방법으로 다층모형이 적합하다고 보았다. 이는 서로 다른 척도의 영향을 제한하기 위해 각 측정치를 표준화하고, 여러 연구의 효과를

비교할 수 있다는 장점이 있다(Wang, Cui, & Parrila, 2011). 또한 다층모형 메타분석은 측정치의 개수가 작을 때에도 데이터 간 자기상관 문제를 고려할 수 있고(Van den Noorgate & Onghena, 2007), 대상 간 측정치의 개수 또는 측정 주기가 다를 때에도 적용 가능하다고 보고된다(Raudenbush & Bryk, 2002). 또한 PND 등 비모수적 방법의 한계를 보완하고자 배빛나, 김소희(2016), 남경욱, 신현기(2008)의 연구에서는 표준화된 효과 차이(standard mean difference, SMD)의 모수적 방법을 함께 사용하여 단일대상 메타분석을 진행한 바 있다.

본 연구에서는 다층모형을 활용하여 단일대상연구의 메타분석을 실시하였고, 구체적으로는 다음의 절차를 거쳤다. 첫째, 개별 연구의 기초학습기능(읽기, 쓰기, 수학)에 해당되는 종속변수 데이터를 코딩하였다. 개별 연구에서 종속변수에 대한 원점수를 명시하지 않고 그래프로만 제시한 경우가 있었는데, 이 경우 그래프 내 데이터 추출을 위한 컴퓨터 프로그램인 GetData Graph Digitizer(2013)를 사용하여 그래프로부터 점수를 추출하였다. 단일대상설계 연구 메타분석에 있어 그래프 데이터 추출을 위한 컴퓨터 프로그램 GetData Graph Digitizer의 신뢰도, 타당도는 상당히 높은 것으로 확인되었다(신미경, 정평강, 2018).

둘째, 본 연구에서는 Van den Noortgate & Onghena (2003a; 2003b; 2007)에서 제시하고 Wang, Cui, & Parrila (2011), Wang, Parrila, & Cui (2013), Heyvaert et al. (2012) 등에서 활용한 방법을 활용하여 데이터를 표준화하고 효과크기를 계산하였다. 이는 원점수를 표준화된 점수로 변환한 뒤 기초선 단계구간의 평균값과 중재 구간의 평균값을 비교하여 효과크기를 계산하는 것이다. 구체적으로는 모든 데이터를 표준화된 점수로 변환하기 위하여 각 연구의 종속변인별 원점수에서 기초선과 중재 구간을 합친 평균 점수를 뺀 뒤 기초선과 중재 구간을 합쳐 계산된 표준편차 값으로 나누었다. 다음으로 각각의 표준화 점수에서 기초선 구간의 표준화점수 평균을 빼주었다. 이는 기초선 단계의 평균을 0으로 조정함으로써 다층모형 분석 시 중재효과에 대한 해석을 용이하게 하기 위함이다. 이러한 방법을 통해 계산된 중재 구간에서의 표준화점수를 종속변인

에 대한 데이터기반교수의 효과로 해석할 수 있다. 다만 이렇게 계산된 단일대상연구별 표준화된 효과크기는 집단설계에서 활용하는 효과크기와 산출방법이 다르기 때문에 동일하게 해석되거나 서로 비교하기는 어렵다는 점을 주의해야 한다(Wang, Cui, & Parrila, 2011).

셋째, 이후 HLM 통계 프로그램을 활용하여 다층메타분석을 실시하였고, 전체 효과크기 및 효과크기의 이질성을 살펴보았다. 단일대상설계 연구의 메타분석에서는 개별 연구가 서로 다른 연구참여자를 포함하고, 개별 연구참여자는 시점에 따른 측정치를 포함하는 위계적 구조를 가진다. 따라서 시점이 개인(연구참여자)에, 개인이 연구에 내재되어 있는 위계적 구조를 지니고, 이러한 자료를 분석하기 위해 이론상으로 3수준 다층모형을 활용하는 것이 적절하다(Van den Noortgate et al., 2013). 그러나 본 연구에서는 단일대상설계 연구의 수가 총 32편으로 3수준 분석을 실시하기에는 충분하지 않았다. 따라서 1수준(시점)과 2수준(연구참여자)으로 구분한 2수준 다층메타분석을 실시하였고, 이를 통해 연구참여자 간 효과크기의 이질성을 설명하는 변인을 탐색하였다.

추가적으로 본 연구에서는 연구참여자별 효과크기를 개선율 차이(IRD) 지수로도 계산하여 부록(<부록 3>)으로 제시하였다. 이는 단일대상 연구의 효과크기 산출을 위해 다양한 효과크기 지수들이 활용될 수 있기에 보다 풍부한 해석을 돕기 위함이다. IRD는 중재 구간에서의 개선율(IR)에서 기초선 구간에서의 개선율(IR)을 뺀 차이를 의미한다. 중재 구간에서의 개선율은 '(기초선의 모든 점보다 높거나 그 이상의 점) / (전체 데이터 점의 수)'로 계산되고, 기초선 구간에서의 개선율은 '(중재 구간보다 높거나 같은 모든 점의 수) / (전체 데이터 점의 수)'로 계산된다(Parker, Vannest, & Brown, 2009; 김동일, 라수현, 이혜은, 2016). IRD는 (－1)에서 (+1)의 범위를 가지는데, 음수는 결과의 악화, 0.5 이하는 '효과가 작거나 없음', 0.5 ~ 0.7 사이는 '중간 효과', 0.7 ~ 0.75 사이는 '높은 효과', 0.75 이상은 '매우 높은 효과'로 해석된다(Parker, Vannest, & Brown, 2009). 연구참여자마다의 독특성을 고려하여, 연구단위로 IRD 평균값을 계산하지 않고, 연구참여자별 IRD 값을 계산하여

제시하였다. 더불어 중재가 여러 구간에 걸쳐 실시되었을 경우, 첫 번째 중재 구간을 기준으로 기초선과 비교하여 IRD를 산출하였다.

한편 본 연구의 데이터 분석 과정에서 추가적으로 고려한 사항은 다음과 같다. 첫째, 단일한 중재가 기초선 및 중재의 두 단계가 아닌 여러 단계에 걸쳐 이루어진 경우, 혼재변수의 잠재적 영향이나 순서에 따른 효과를 최소화하기 위하여 가장 처음의 중재 단계의 데이터를 코딩하였다. 분석대상논문에는 중 기초선 구간과 중재 구간으로 구분하지 않고 전체 학급 대상 중재(RTI 1단계)와 소집단 중재(RTI 2단계)로 구분한 경우도 있었는데, 이 경우에는 전자를 기초선 구간, 후자를 중재 구간으로 코딩하였다. 둘째, 모든 효과크기는 양수의 값이 수행 수준의 진전을 나타낼 수 있도록 하였다. 즉 오류율을 측정할 경우 음수의 효과크기가 나타나는데 이에 (-1)을 곱하여 양수의 효과크기를 가질 수 있도록 조정하였다. 셋째, 대부분의 분석대상논문에서 복수의 종속변인을 설정하고 있었는데, 그 중 기초학습능력에 해당되는 것만 코딩하였다. 넷째, 본 연구에서는 한 편의 연구 및 한 명의 연구참여자로부터 여러 개의 효과크기가 산출되는 것을 지양하였다. 이는 동일한 대상 또는 연구에서 여러 개의 효과크기가 도출될 경우 효과크기가 서로 독립적이라고 보기 어렵기 때문이다(Hedges, Tipton, & Johnson, 2010). 따라서 본 연구에서는 읽기, 쓰기, 수학 등 각 기초학습능력영역 안에서 복수의 측정치가 제기되어 있을 경우 대표적인 종속변인 측정치를 채택하였다. 구체적으로는 읽기유창성을 속도, 정확성의 두 가지 기준으로 측정할 경우 속도를, 쓰기능력을 유창성과 정성적 점수로 측정할 경우 쓰기유창성을 코딩하였다. 또한 쓰기를 완전한 문장수, 어휘수 등으로 측정할 경우 완전한 문장수의 데이터를 코딩하였고, 종속변인을 BASA 초기수학 검사와 텍사스 조기수학검사로 측정하였을 경우 BASA 검사의 결과값을 활용하였다. 그리고 동일한 척도의 검사에서 점수를 영역별로 제시하고 있는 경우에는 영역별 원점수의 평균 값을 계산하여 활용하였다. 다만 읽기와 쓰기 등 서로 다른 기초학습능력 영역을 동시에 종속변인으로 설정하고 있는 경우, 한 개만 택하지 않고 복수의 효과크기를 독립적으로 사용하였다.

Ⅳ. 연구결과

1. 분석대상논문의 특성

총 32편의 단일대상연구가 분석대상으로 선정되었다. 32편의 연구에는 총 100명의 연구참여자가 포함되어 있었고, 전체 연구참여자에 걸쳐 기초학습영역(읽기, 쓰기, 수학)에서 총 1,601개의 시점에 따른 측정치와 총 119개의 효과크기가 산출되었다. 이를 요약하면 <표 Ⅳ-1>과 같다.

<표 Ⅳ-1> 단일대상설계 연구의 개요

저자	학생 수(남, 여)	효과크기 수
심말옥, 박현옥(2015)	3(0,3)	6
권혜정(2014)	3(1,2)	3
조경화(2008)	3(3,0)	6
최종근(2008)	1(0,1)	1
유은경(2016)	3(2,1)	3
김경희(2017)	3(3,0)	3
이혜진, 이태수(2019)	3(1,2)	3
권덕용, 강옥려(2007)	3(3,0)	3
유종완(2010)	3	3
한연숙(2014)	3(1,2)	3
박수현, 이기정(2015)	3(1,2)	6
이은옥(2015)	3(3,0)	3
권미영(2015)	3(2,1)	3
신수정, 강옥려(2018)	3(2,1)	9
김동일, 고혜정, 이해린(2014)	2(1,1)	2

김보람(2011)	6	6
고영실, 송재홍(2019)	1(0,1)	1
김미진, 이재원, 이동철(2014)	2(1,1)	2
이진영(2015)	3(3,0)	3
김민경, 이기정(2015)	3(2,1)	3
길유미(2015)	3(3,0)	3
설영기(2005)	3(2,1)	3
김동일 외(2018)	5(3,2)	5
황애희(2012)	5(2,3)	5
안성진(2018)	4(2,2)	8
최은주(2006)	4(3,1)	4
정광조, 이효자(2009)	4(2,2)	4
송푸름, 김동일(2020)	3(1,2)	3
김은삼(2015)	3(2,1)	3
안미선(2012)	3(3,0)	3
정채미(2015)	3(2,1)	3
김동일 외(2020)	3(2,1)	3
합계	100	119

분석대상논문의 특징을 확인하기 위해 일반적인 특징, 연구참여자 관련 특징, 중재 관련 특징, 데이터기반교수 관련 특징으로 구분하여 기술통계를 제시하면 다음과 같다.

먼저 <표 Ⅳ-2>에서 확인할 수 있듯이 학령기 학습곤란 학생을 대상으로 데이터기반교수를 적용한 단일대상설계 연구의 수는 2005년을 시작으로 점차 증가하는 경향을 보였고, 2015년 이후 현재까지 활발하게 연구되었다. 출판유형별로 구분하였을 때 학술지보다 학위논문이 다소 더 많은 비중을 차지하였으며, 연구설계별로 살펴보면 중다기초선설계가 12

편으로 가장 높은 비율을 차지하였다. 단일대상설계 연구 중 구체적인 연구설계방법을 밝히지 않았으나, 중재를 실시하지 않은 구간과 실시한 구간이 구분되어 있어 기초선과 중재 구간으로 간주할 수 있으며, 연구참여자의 종속변수 변화를 개별적으로 제시하고 있는 10편의 연구는 기타로 분류되었다.

<표 Ⅳ-2> 단일대상설계 연구의 일반적인 특징

구분	연구 수	비율
출판연도		
2005-2009	6	18.8
2010-2014	8	25.0
2015-2020	18	56.3
출판유형		
학술지	14	43.8
학위논문	18	56.3
연구설계		
중다기초선설계	12	37.5
AB 설계	5	15.6
ABA 설계	3	9.4
ABAB 설계	1	3.1
교대 중재 설계	1	3.1
기타	10	31.3
합계	32	100

다음으로 <표 Ⅳ-3>의 연구참여자 관련 특징을 살펴보면, 32편 중 28편이 초등학생을 대상으로 하고 있었고, 그 중 특히 1~3학년의 초등학교 저학년을 대상으로 하는 경우가 절반 이상이었다. 중학교와 고등학교 학습곤란 학생을 대상으로 데이터기반교수를 적용한 연구는 각각 3편, 1편으로 드문 편이었다. 다음으로 개별 논문에서 사용한 용어를 바탕으로 연구에 참여한 학생 유형을 구분하여 보면, 학습부진 학생을 대상으로 한 경우가 총 14편(43.8%)으로 가장 많았다. 이어서 학습장애 및 학습장

애 위험 학생(28.1%), 지적장애 학생(15.6%), 저성취 학생(6.3%) 순으로 많았으며, 이 외에도 기타로 분류된 2편의 연구에서는 각각 난독증 위험과 ADHD 학생을 대상으로 하고 있었다. 다만 각각의 연구에서 참여자 선정을 위해 사용한 기준을 검토한 결과, 저성취와 학습부진 용어는 명확하게 구분되지 않고 있었다. 학교 유형의 경우, 절반 이상이 일반학급에 해당되었고, 특수학급과 특수학교는 각각 6편, 1편으로 적었다. 혼합으로 분류된 3편의 연구 중에는 일반학급과 특수학급 학생이 함께 포함된 연구, 일반학급과 지역아동센터 학생이 함께 포함된 연구가 각각 2편과 1편씩 있었다. 이 외 아동복지시설 및 지역아동센터 학생만을 대상으로 한 경우는 기타로 분류하였다.

<표 Ⅳ-3> 단일대상설계 연구의 연구참여자 관련 특징

구분		연구 수	비율	
학교급				
초등	초저(1-3)	19	59.4	87.6
	초고(4-6)	7	21.9	
	초저+초고	2	6.3	
중등		3	9.4	
고등		1	3.1	
학생유형				
학습장애 진단/위험		9	28.1	
저성취		2	6.3	
학습부진		14	43.8	
지적장애		5	15.6	
기타		2	6.3	
학교 유형				
일반학급		20	62.5	
특수학급		6	18.8	
특수학교		1	3.1	
혼합		3	9.4	
기타		2	6.3	
합계		32	100	

다음으로 분석대상논문의 중재 관련 특징은 <표 Ⅳ-4>와 같다. 전체적으로 총 회기 수는 10~19회기(71.9%), 중재 기간은 4주~9주(68.8%)가 가장 많음을 확인할 수 있었다. 주 2회로 실시된 경우가 25편으로 대부분이었으며, 회기 당 시간은 31~45분 범위에 있는 경우가 23편으로 가장 많았다. 또한 중재가 이루어진 교수집단의 크기를 살펴보면, 일대일로 이뤄진 경우가 23편으로 가장 많았는데, 이는 단일대상설계의 특징에 더불어 개별화된 접근으로 활용되는 데이터기반교수의 특징과도 관련이 있다고 볼 수 있다. 중재 실시자의 경우 연구자와 동일한 경우가 24편으로 대부분이었지만, 4편의 연구에서는 누가 중재를 제공했는지 정보를 제공하지 않고 있었다. 중재 장소의 경우 일반학급에서 중재를 제공한 연구가 절반이었으며, 특수학급, 특수학교 등에서 실시한 경우는 비교적 적었다. 중재 장소로서 일반학급과 지역복지시설을 함께 사용한 경우, 학습도움실이나 교사연구실과 같은 독립된 공간을 사용한 경우는 기타로 분류하였다. 다만 중재 장소에 대해서도 6편의 연구에서는 정보를 제시하지 않고 있었다. 이어서 학업 태도, 학업 관련 자아개념, 문제행동 등을 제외하고 기초학습영역에서 중재의 목표가 된 종속변인을 중심으로 살펴본 결과, 전체 연구의 절반에서 읽기를 종속변인으로 하고 있었다. 특히 그 중 읽기유창성을 목표 기술로 한 경우와 읽기유창성과 읽기이해를 함께 목표로 한 연구가 각각 7편으로 많았다. 음운인식 및 단어재인 등 초기문해를 단독으로 종속변인으로 설정한 경우는 1편에 불과했다.

<표 Ⅳ-4> 단일대상설계 연구의 중재 관련 특징

구분	연구 수	비율
중재 회기 수		
10-19회기	23	71.9
20-29회기	7	21.9
30회기 이상	2	6.3
주당 회기 수		
주 1회	5	15.6
주 2회	25	78.1

주 3회	2	6.3	
회기 당 시간			
15-30분	2	6.3	
31-45분	23	71.9	
46-60분	6	18.8	
미제시	1	3.1	
중재 기간			
4-9주	22	68.8	
10-14주	9	28.1	
15-19주	1	3.1	
집단 크기			
일대일	23	71.9	
소집단(2-3명)	3	9.4	
중집단(4-6명)	2	6.3	
일대일+집단	2	6.3	
미제시	2	6.3	
중재 실시자			
연구자	24	75.0	
대학원생	2	6.3	
연구자+자원봉사자	1	3.1	
또래	1	3.1	
미제시	4	12.5	
중재장소			
일반학급	16	50.0	
특수학급	3	9.4	
특수학교	1	3.1	
치료실(사설센터)	1	3.1	
기타	5	15.6	
미제시	6	18.8	
종속변인 영역			
읽기 초기문해	1	3.1	50.0

	읽기유창성	7	21.9	
	읽기유창성+읽기이해	7	21.9	
	초기문해+유창성+읽기이해	1	3.1	
쓰기		6	18.8	
수학	초기수학	1	3.1	25.0
	연산	7	21.9	
읽기+쓰기		2	6.3	
합계		32	100	

마지막으로 32편의 단일대상연구에서 나타난 데이터기반교수 관련 특징을 분석한 결과는 <표 Ⅳ-5>과 같다. 먼저 진전도 모니터링을 위해 사용된 CBM 측정 도구의 유형을 살펴보면, 읽기유창성 검사(50%), 수학 검사(21.9%), 쓰기 검사(18.8%) 순으로 높은 비율로 활용됨을 알 수 있었다. 학생의 성장을 점검하기 위한 진전도 측정 주기를 살펴보면, 매회기 측정을 실시한 경우가 20편으로 가장 많았고, 이어서 2~3회기 마다 측정한 경우가 9편으로 나타났다. 이렇듯 전반적으로 진전도 측정이 빈번하게 이루어진 것은 측정이 효율적이고 간편한 CBM의 특징과도 관련이 있다고 볼 수 있다. 진전도 측정 실시자는 대부분 중재자였지만, 이를 밝히지 않은 연구도 6편 있었다.

한편 학생의 진전도 데이터를 바탕으로 어떤 기준에 의해 교수적 의사결정 또는 수정을 하였는지를 나타내는 데이터기반 의사결정 규칙의 경우, 13편에 달하는 연구에서 충분한 정보를 제시하지 않고 있었다. 관련 정보를 제시한 연구 중에는 BASA 검사의 지침서를 바탕으로 기울기 방법과 데이터포인트 방법을 혼합하여 활용한 경우와 숙달기준 방법을 단독으로 사용한 경우가 각각 7편, 6편으로 높은 비중을 차지하였다. 일반적으로 증거기반 의사결정 규칙 유형으로 제시되지는 않지만, 또래 규준에 비추어본 백분위 점수를 활용한 경우도 3편 있었다. 다만 기울기 방법과 달리 교수적 의사결정을 위하여 데이터포인트 방법만을 단독으로 사용한 연구는 분석대상논문 중 없는 것으로 나타났다.

마지막으로 교수적 수정 유형을 자료 및 과제의 난이도, 개별화된 접

근, 교수전략 등을 조정하는 질적 수정과 교수 집단의 크기, 중재 빈도 및 총 회기 수 등을 조정하는 양적 수정으로 구분하여 살펴보았다. 분석 결과, 22편에 달하는 연구에서 CBM을 통한 진전도 측정을 수행했음에도 불구하고, 측정 데이터를 바탕으로 어떠한 교수적 수정을 가하였는지에 대한 충분한 정보를 제공하지 않은 것으로 나타났다. 나머지 10편의 연구 중에는 질적 수정을 실시한 연구가 6편, 양적 수정과 질적 수정을 함께 실시한 연구가 3편이었고, 질적 측면에서의 조정 없이 양적 수정만을 가한 연구는 1편에 불과하였다.

<표 Ⅳ-5> 단일대상설계 연구의 데이터기반교수 관련 특징

구분	연구 수	비율
CBM 유형 (BASA 검사)		
초기문해	1	3.1
읽기유창성	16	50.0
읽기유창성+읽기이해	1	3.1
쓰기	6	18.8
초기수학	1	3.1
수학	7	21.9
진전도 측정 주기		
매회기	20	62.5
2-3회기마다	9	28.1
4회기마다	2	6.3
변동적	1	3.1
진전도 측정 실시자		
중재자	24	75.0
중재자 이외	2	6.3
미제시	6	18.8
증거기반 의사결정 규칙		

기울기	1	3.1
숙달기준	6	18.8
기울기+데이터포인트	7	21.9
기울기+숙달기준	1	3.1
백분위	3	9.4
기타	1	3.1
미제시	13	40.6
교수적 수정 유형		
양적	1	3.1
질적	6	18.8
양적+질적	3	9.4
미제시	22	68.8
합계	32	100

2. 질적 지표 분석 결과

학령기 학습곤란 학생 대상 데이터기반교수를 적용한 32편의 분석대상논문이 증거기반실제로서 수용 가능한 정도를 확인하고자 질적 분석을 하였다. Horner 외(2005)가 제시한 단일대상연구 질적 지표를 바탕으로 하되, 엄격한 기준 및 이분법적 평가의 제한점 등을 고려하여 질적 분석을 실시한 선행연구(이예다나 외, 2015; 안예지, 이미지, 2020; 최희승, 2020)를 참고하여 분석하였다. 분석은 3점 평가척도로 이루어졌고, 마찬가지로 2점 이상의 평균을 보였을 때 충분한 질적 수준을 갖춘 것으로 해석하였다.

질적 지표 분석 결과를 요약하여 제시하면 <표 Ⅳ-6>과 같다. 총 32편의 연구 중 30편의 연구에서 전체 지표 평균 2점을 넘어 충분한 질적 수준을 갖추고 있음을 확인할 수 있었다. 특히 종속변수의 양적 측정, 3

번 이상의 효과치를 통한 실험 효과 입증, 패턴 및 수준 변화를 통한 결과의 증명, 종속변수의 사회적 중요성 지표의 경우 32편의 분석대상논문 모두에서 3점으로 나타났다.

평균 점수를 중심으로 질적 지표별 분석결과를 구체적으로 살펴보면 다음과 같다. 먼저 연구대상자 및 연구환경 영역에서는 연구대상자 기술, 연구대상자 선별과 관련하여 평균 점수가 각각 2.88점, 2.78점으로 대부분의 연구에서 높은 질적 수준을 갖추고 있으나, 중재제공자 기술과 중재가 제공된 물리적 환경에 대한 기술에서는 평균 2점 이하로 충분한 기술이 이루어지지 않은 연구가 많았다.

다음으로 종속변수 영역에서는 종속변수 기술, 측정과정 기술, 측정의 타당도, 측정 빈도 관련 지표에서 점수 평균이 2.88점에서 3점으로 매우 높았고, 모든 연구에서 충분한 질적 수준을 갖추고 있는 것으로 나타났다. 측정 신뢰도의 경우 평균이 1.63점으로 비교적 낮았는데, 이는 22편의 연구에서 측정 신뢰도 또는 관찰자간 일치도 자료를 수집하지 않거나 보고하지 않았기 때문이다. 다만 이는 대부분의 연구에서 표준화된 검사도구를 통해 객관적으로 종속변수를 측정하고 있어, 측정기준에 대한 측정자 내의 일관성 있는 평가 또는 수행을 평가하는 관찰자간 일치도의 측정으로부터 비교적 자유로웠을 것으로 보인다.

독립변수 영역의 경우, 독립변수 기술에서는 모든 연구에서 충분히 기술하고 있어 평균 2.75점으로 높게 나타났고, 구체적으로는 24편의 연구에서 연구가 반복 가능하도록 독립변인을 정확하게 기술하고 중재의 개념적 기초, 중재 회기 수 및 기간, 세부적인 교수 과정 및 교수 자료 등에 대한 정보를 자세히 제공하는 것으로 나타났다. 독립변수 조작과 중재 충실도 지표에서는 각각 1.91점, 1.59점으로 충분한 질적 수준을 확보하였다고 보기 어려웠다. 이는 연구자의 통제 하에서 언제 중재를 투입하고, 단계에 변화를 주며, 중재를 철회하는지 등 독립변수 조작 및 실험 통제에 대한 정보를 구체적으로 제시하지 않은 연구가 14편으로 상당히 많았기 때문이다. 또한 22편의 연구에서 중재 충실도를 직접적으로 측정하지 않고 있어 질적 제고가 필요한 것으로 분석되었다.

기초선 영역의 종속변수 측정 지표에서는 평균 점수 2.47점으로, 3편을 제외한 연구에서 기초선 기간 동안 종속변수를 3회 이상 측정하였고 중재 실시 전 안정화됨을 확인하였다. 기초선 기간에 대한 기술은 평균 1.75점으로 비교적 낮게 나타났으나, 20편의 연구에서는 2점 이상으로 충분한 질적 수준을 갖추고 있었다.

실험통제 및 내적타당도 영역에서는 전반적으로 높은 질적 수준을 확보하고 있음을 확인하였는데, 32편의 연구에서 모두 3번 이상의 측정치에서 실험 효과를 입증하고 있었고, 그래프의 시각적 분석을 통하여 경향 또는 수준 변화를 밝힘으로써 결과 패턴을 보고하고 있었다. 다만 내적타당도 지표에서는 평균 2.03점으로 충족 기준은 넘었으나, 12편의 연구에서 시간적 흐름, 외부적 사건, 측정의 효과, 측정도구의 효과, 실험도중 탈락 발생 등 내적 타당도에의 위협요소를 어떻게 통제하였는지 구체적으로 서술하지 않고 있었다. 이에 질적 수준 제고를 위한 노력이 보다 필요하다고 볼 수 있다. 또한 구체적인 연구 설계에 따라 내적타당도 지표의 점수 편차가 있었는데, 중다기초선, 중다간헐기초선 설계를 활용한 연구에서는 비교적 높은 점수를 획득한 반면, 구체적인 연구 설계를 밝히지 않은 단일대상 및 사례 연구에서는 전반적으로 낮은 점수를 획득한 경향이 있었다.

다음으로 외적 타당도 영역의 경우 평균 2.53점으로 충분한 질적 수준을 충족하고 있었다. 구체적으로 2편을 제외한 연구에서 대상자 간 또는 검사도구 간 중재 효과가 반복됨을 확인하고 있었다.

마지막으로 사회적 타당도 영역에서는 지표별로 1.47점에서 3점으로 지표별 평균점수 편차가 큰 편이었다. 종속변인의 사회적 중요성에서는 평균 3점으로 모든 연구에서 기초학습능력 또는 학습 관련 태도 등 사회적 유의미성을 가진 종속변수를 설정하고 있음을 확인할 수 있었다. 종속변수 변화 크기의 사회적 중요성 역시 평균 2.63점으로 높았는데, 구체적으로 대부분의 연구에서 평균 수준, PND, IRD 등으로 변화 정도 및 효과크기를 기술하고 있었고, 그 변화 크기가 사회적 중요성을 지니는 것으로 판단되었다. 다만 독립변수의 수행 실용성 및 비용 효과성, 독립

변수의 수행 특성과 관련된 지표는 충분한 질적 수준에 미치지 못했고, 전체 질적 지표 중 평균 점수가 가장 낮은 수준이었다. 구체적으로 학생, 학부모 등에게 사회적 타당도 자료를 수집하거나 수용성, 중재 적용 가능성, 효과성, 계속적인 사용 등에 대한 정보를 제시한 연구는 5편에 불과하여 질적 개선이 필요한 것으로 나타났다. 또 연구에서 설정한 중재 상황이 아닌 일반적인 상황 속 일반적인 중재자에 의해서도 독립변수의 효과를 확인할 수 있도록 독립적인 수행에 대한 특성을 충분히 기술한 연구는 한 편에 불과하였다.

<표 Ⅳ-6> 단일대상설계 연구의 질적 지표 분석 결과

분류	구체적 지표	논문 수			평균
		1 (미충족)	2 (부분충족)	3 (충족)	
연구대상자 및 연구환경	연구대상에 대한 정보를 구체적으로 서술하였는가? (예: 나이, 성별, IQ, 장애유무, 진단명, 사회경제적 수준, 학습수준 평가결과 등)	0	4	28	2.88
	연구대상자 선별기준 및 선정절차가 반복 가능하도록 정확하게 제시되었는가?	1	5	26	2.78
	중재제공자에 대한 정보를 구체적으로 서술하거나 (예: 경력, 교사자격증, 교육수준, 나이, 성별, 인종, 중재친숙도 등), 조건 간 중재제공자의 동등성에 대해 기술하였는가?	21	7	4	1.47
	연구가 수행된 물리적 환경이 구체적으로 기술되었는가?	9	17	6	1.91
종속변수	종속변수를 조작적으로 명확하게 기술하고 있는가?	0	1	31	2.97
	모든 종속변수에 대하여 양적으로	0	0	32	3.00

	측정가능한 지표를 산출하도록 측정되었는가?				
	측정이 타당하며 반복 가능하도록 측정과정에 대해 명확하게 기술하였는가?	0	4	28	2.88
	측정이 충분한 빈도수로 반복적으로 수행되었는가?	0	3	29	2.91
	각각의 종속변인에 대해 측정 신뢰도 또는 관찰자간 일치도 자료가 수집되었으며, 최소한의 기준에 도달하였는가?	22	0	10	1.63
독립변수	독립변인이 반복 가능하도록 정확하게 기술되었으며 자세한 정보를 제공하였는가? (예: 중재의 개념적 기초, 교수 자료 및 과정, 중재기간, 회기 시간, 교사의 행동 및 언어 등)	0	8	24	2.75
	연구자 통제 하 독립변수가 조작되고 실험통제에 대해 정확하게 기술하였는가?	14	7	11	1.91
	중재 충실도에 대한 측정과 명확한 기술이 이루어졌는가?	22	1	9	1.59
기초선	기초선 기간에 종속변수가 자주 측정되었고 앞으로의 수행을 예측할 수 있도록 중재 전 안정화되었는가?	3	11	18	2.47
	기초선 기간에 대해 반복 가능하도록 구체적인 정보를 제시하였는가? (예: 절차, 환경 등)	12	16	4	1.75
실험통제/내적 타당도	실험효과가 3번 이상의 측정치에서 충분히 증명되었는가?	0	0	32	3.00
	내적 타당도 위험요소를 대부분 통제하였는가?	12	7	13	2.03

	결과가 증명되었는가? (예: 결과 또는 수준의 변화)	0	0	32	3.00
외적 타당도	독립변인의 효과가 대상자 간, 환경 간, 검사도구 간 반복되었는가?	2	11	19	2.53
사회적 타당도	종속변수가 사회적으로 중요한가?	0	0	32	3.00
	중재를 통한 종속변수의 변화 정도가 사회적으로 중요한가? (예: 평균 수준, PND 등)	0	12	20	2.63
	독립변인 실행의 실용성, 비용효과성에 대한 설명을 제공하였는가? (예: 사회적 타당도 자료 제공, 수용성, 중재적용가능성, 효과성과 계속적인 사용에 대한 설명)	22	5	5	1.47
	독립변수의 독립적인 수행 특성에 대한 정보를 제공하였으며. 일반적인 상황 속 일반적인 중재자에 의해서도 향상 가능한가?	16	15	1	1.53

3. 효과크기 분석 결과

1) 모형 설정

HLM 프로그램으로 2수준 다층메타분석을 통해 학습곤란 학생 대상 데이터기반교수의 평균 효과 크기와 잠재적인 예측변인(학생 관련, 중재 관련, 데이터기반교수 관련 변인)이 효과크기에 미치는 영향을 분석하였다. 단일대상설계의 시점-개인-연구의 위계적 구조를 고려할 때 3수준 다층메타분석이 적절하나, 분석대상논문 수가 3수준 모형을 적용하기에는 충분치 않아 2수준 다층모형을 적용하였다. 1수준은 연구참여자별 시

점 측정치, 2수준은 전체 연구참여자로 설정한 뒤 분석을 실시하였다.

1수준 (시점 수준):

$$Y_{ij} = \pi_{1j}(PHASE) + e_{ij} \qquad e_{ij} \sim N(0, \sigma^2)$$

2수준 (연구참여자 수준):

$$\pi_{1j} = \beta_{10} + r_{1j} \qquad r_{1j} \sim N(0, \tau_{oo}^2)$$

먼저 본 연구에서는 연구참여자의 전체 효과크기를 확인하고, 중재 효과의 연구참여자 간 차이를 확인하기 위해 변형된 기초모형을 위와 같이 설정하였다.

1수준 모형에서 Y_{ij}는 j 번째 참여자의 i 번째 시점에서의 표준화 측정치를 의미한다. 본 연구는 중재의 효과를 확인하고 종합하는 데 목적이 있으므로 기초모형에 $PHASE$ 변수를 투입하였고, 기초선 구간 평균을 0으로 조정한 상태이므로 절편을 생략하였다. $PHASE$ 변수는 기초선 구간을 0, 중재 구간을 1로 더미코딩한 변수로, 변수를 투입함으로써 기초선 구간에 비해 중재가 투입된 이후 기초학습능력의 표준화 측정치가 얼마나 향상하였는지를 알 수 있다.

1수준 모형에서 π_{1j}는 j 번째 참여자가 중재를 제공받음에 따라 기초선 구간에 비해 중재 구간에서 변화된 정도를 의미한다. 이는 곧 중재 효과로 해석할 수 있으며, 표준화된 측정치로 변환한 뒤 분석하였기에 표준화된 중재 효과로 해석 가능하다. 다음으로 2수준 모형에서 β_{10}는 π_{1j}의 평균으로, 전체 평균 효과크기를 나타낸다. 즉 32편의 분석대상논문에 포함된 연구참여자 100명의 평균 중재 효과크기를 의미하고, 다시 말해 데이터기반교수 원리를 따른 중재가 투입되었을 때 연구에 참여한 학생들의 기초학습능력은 평균적으로 얼마나 증가했는지를 나타낸다.

1수준 (시점 수준):

$$Y_{ij} = \pi_{1j}(PHASE) + e_{ij} \qquad e_{ij} \sim N(0, \sigma^2)$$

2수준 (연구대상 수준):

$\pi_{1j} = \beta_{10} + \beta_{11}(\text{학생유형}) + r_{1j} \qquad r_{1j} \sim N(0, \tau_{oo}^2)$

다음으로 중재 효과크기에 영향을 미치는 변인을 탐색하기 위하여 2수준 모형의 π_{1j}에 예측변인을 투입한 조건모형을 설정하였다. 예측변인 투입의 목적은 효과크기의 연구참여자 간 분산을 설명하는 변수를 탐색하기 위함이므로 중재효과를 나타내는 π_{1j}에 대한 모형에 변수를 투입하였다. 예측변인들을 함께 모형에 투입할 경우 다른 변수를 통제한 후의 특정 변수의 영향을 확인할 수 있는 장점이 있는 반면, 한꺼번에 많은 수의 예측변인을 투입함으로써 다중공선성의 위험이 증가하여 해석이 잘못될 가능성이 있다(Heyvaert et al., 2012). 따라서 본 연구에서는 예측변인을 조건모형에 한 개씩 개별적으로 투입하여 그 영향을 확인하였다. 예측변인의 조절효과 확인을 위해 활용된 조건모형의 예시는 위와 같고, 해당 모형에서는 예측변인으로 학생유형 변인을 투입한 상태이다.

본 연구에서 적용한 분석 모형을 정리하면 <표 Ⅳ-7>과 같다.

<표 Ⅳ-7> 단계별 분석 모형

모형	투입변수	목적
모형 1 (변형된 무조건모형)	중재단계 (PHASE)	전체 효과크기 확인 및 연구참여자 간 효과크기의 이질성 분석
모형 2 (조건모형)	중재단계 (PHASE) + 예측변인	연구참여자 수준에서 효과크기의 이질성에 영향을 미치는 변수 확인

2) 동질성 검증 및 평균 효과크기 산출

먼저 무조건모형의 분석결과를 살펴보면 <표 Ⅳ-8>과 같다. 분석 결

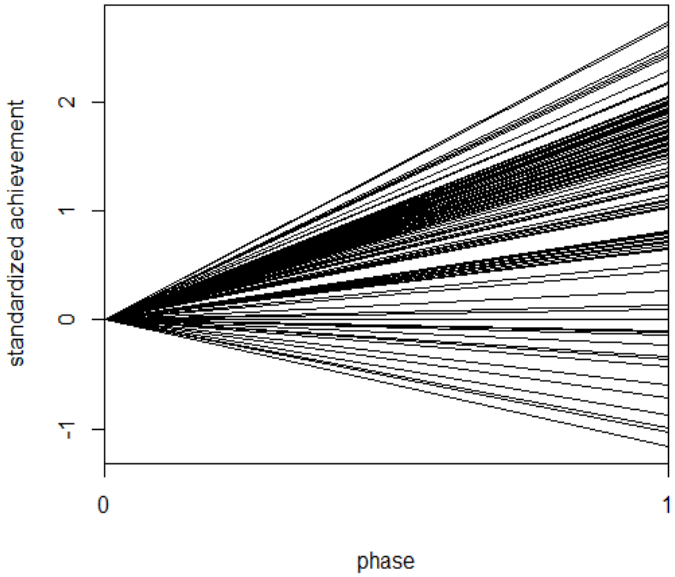
과, 전체 연구참여자에 대한 평균 중재 효과, 즉 기울기 계수는 1.341이고, 통계적으로 유의했다. 기울기 계수의 신뢰도는 0.862로 높았다. 즉 학령기 학습곤란 학생 대상 데이터기반교수 원리를 적용한 중재의 효과는 0이 아니며, 학생들의 기초학습능력은 평균적으로 1.341 표준편차만큼 증가했다고 해석할 수 있다. 또한 무선효과 분석 결과에서 중재 효과의 표준편차는 0.722이었으며, 중재 효과의 무선효과가 통계적으로 유의했다. 이는 중재 효과의 연구참여자 간 차이가 존재함을 의미하고, 따라서 이러한 효과의 차이를 설명하기 위한 변인의 탐색이 필요성이 있음을 알 수 있다.

<표 Ⅳ-8> 무조건 모형 분석 결과

고정효과	회귀계수	표준오차	t	자유도	유의확률
기울기(β_{10})	1.341	0.078	17.224***	99	<0.001
무선효과	**표준편차**	**분산성분**	χ^2	**자유도**	**유의확률**
기울기(r_{1j})	0.722	0.522	907.154***	99	<0.001
1수준 오차	0.852	0.725			

*p<.05 **p<.01 ***p<.001*

이러한 분석 결과는 연구참여자별 효과크기를 표현한 그림을 통해서도 이해할 수 있다. [그림 Ⅳ-1]은 Van den Noortgate & Onghena (2003a; 2003b; 2007), Wang, Cui, & Parrila (2011), Wang, Parrila, & Cui (2013)에서 사용된 방법으로 개별 연구참여자의 표준화된 효과크기를 산출한 뒤 시각적으로 나타낸 것이다. x축의 0은 기초선 구간, 1은 중재 구간을 의미하고, y축은 기초학습능력의 표준화된 측정치를 의미한다. 직선의 기울기는 각각 100명의 연구참여자에서 산출된 119개의 표준화된 효과크기에 해당된다. 무선효과 분석 결과에서 확인한 바와 같이, 시각적으로도 연구참여자별 효과크기가 상당히 넓은 범위에 걸쳐 퍼져있고, 이에 예측변인 탐색의 필요성을 확인할 수 있다.



[그림 Ⅳ-1] 연구참여자별 효과크기에 대한 시각적 분석

3) 기초학습영역별 평균 효과크기 산출

추가적으로 데이터기반교수의 평균 효과크기를 기초학습영역별로 확인하고자, 종속변인을 읽기, 쓰기, 수학의 세 영역으로 구분하여 분석하였다. 읽기 영역에는 음운인식, 단어재인, 읽기유창성, 읽기이해 능력을 측정한 연구들이, 쓰기 영역에는 쓰기 능력의 정성적, 정량적 점수를 측면을 측정한 연구들이, 수학 영역에는 수감각, 수세기 등 초기수학과 연산능력을 측정한 연구들이 포함되었다.

영역별 분석결과는 <표 Ⅳ-9>와 같다. 가장 많은 비중을 차지한 읽기 영역의 경우 60명의 연구참여자로부터 916개의 시점 측정치가 산출되었다. 평균 중재효과는 1.432로 통계적으로 유의했다. 쓰기 영역에서는 24명의 연구참여자와 586개의 시점 측정치가 분석에 활용되었고 평균 효과크기는 1.520으로 통계적으로 유의했으며, 세 영역 중 효과가 가장 높은 것으로 나타났다. 수학 영역의 경우 22명의 연구참여자와 275개의 시점

측정치가 분석되었다. 수학 영역의 평균 중재 효과는 1.091로 통계적으로 유의했고, 읽기와 쓰기에 비해 상대적으로 낮았으나 여전히 높은 수준의 효과크기에 해당되었다. 연구참여자에 따른 중재 효과의 이질성을 확인한 결과, 읽기 영역에서는 무선효과가 유의하지 않았으나, 쓰기와 수학 영역에서는 무선효과가 통계적으로 유의하여 효과크기가 연구참여자의 특성에 따라 상이함을 알 수 있다.

<표 Ⅳ-9> 기초학습영역별 무조건 모형 분석 결과

읽기					
고정효과	**회귀계수**	**표준오차**	**t**	**자유도**	**유의확률**
기울기(β_{10})	1.432	0.090	15.939***	59	<0.001
무선효과	**표준편차**	**분산성분**	χ^2	**자유도**	**유의확률**
기울기(r_{1j})	0.633	0.400	429.965	59	>0.500
1수준 오차	0.833	0.694			
쓰기					
고정효과	**회귀계수**	**표준오차**	**t**	**자유도**	**유의확률**
기울기(β_{10})	1.520	0.165	9.188***	23	<0.001
무선효과	**표준편차**	**분산성분**	χ^2	**자유도**	**유의확률**
기울기(r_{1j})	0.782	0.611	370.814***	23	<0.001
1수준 오차	0.802	0.643			
수학					
고정효과	**회귀계수**	**표준오차**	**t**	**자유도**	**유의확률**
기울기(β_{10})	1.091	0.188	5.815***	21	<0.001
무선효과	**표준편차**	**분산성분**	χ^2	**자유도**	**유의확률**
기울기(r_{1j})	0.823	0.678	208.508***	21	<0.001
1수준 오차	0.847	0.718			

*p<.05 **p<.01 ***p<.001

4) 예측변인의 조절효과 분석 결과

변형된 기초모형의 무선효과 분석 결과, 데이터기반교수의 효과크기가 연구참여자에 따라 동질적이지 않음을 확인하였다. 이에 2수준 중재효과(π_{1j}) 모형에 대하여 예측변인을 투입한 조건모형 분석을 통하여 효과크기의 이질성에 영향을 미치는 변인을 탐색하였다. 중재효과의 이질성을 설명하기 위해 조건모형에 투입된 변인들과 코딩사항은 <표 Ⅳ-10>과 같다.

모든 변수는 분석을 위하여 더미코딩되었는데, 분석이 가능한 충분한 사례 수를 고려하여 두 개 집단으로 구분하여 코딩하였다. 분석에 있어 특정 변인에 대한 정보를 정확히 제시하지 않은 경우에는 목록 삭제(listwise) 방식으로 결측 처리하여 해당 연구참여자의 사례가 모두 분석에서 제외되도록 하였다. 학생유형 변수의 경우 학습장애 진단 및 위험, 저성취 및 학습부진, 지적장애 각각에 대하여 해당 여부를 더미코딩한 변수를 모형에 투입하였다.

<표 Ⅳ-10> 조절효과 분석에 투입된 예측변인

예측변인	코딩 사항	
	0	1
학생 관련		
학교급	초등	중고등
학생유형	학습장애 진단/위험	학습장애 진단/위험 아님
	저성취 및 학습부진	저성취 및 학습부진 아님
	지적장애	지적장애 아님
성별	남자	여자
학교유형	일반학급	일반학급 외

중재 관련		
전체 회기 수	19회기 이하	19회기 초과
주당 회기수	1회	2-3회
1회기 당 중재 시간	45분 이하	45분 초과
중재 기간	9주 이하	9주 초과
집단 크기	일대일	2명 이상
데이터기반교수 관련		
진전도 측정 주기	매회기	2회기 이상마다
증거기반 의사결정 유형	기울기 방법을 활용함	기울기 방법을 활용하지 않음
교수적 수정 내용	제시함	제시하지 않음

조건모형 분석 결과를 학생 관련, 중재 관련, 데이터기반교수 관련 변인으로 구분하여 제시하면 같다. 첫째, 학생 관련 변인을 투입한 조건모형 분석 결과는 <표 Ⅳ-11>과 같다. 회귀계수가 유의하다는 것은 해당 변인의 구분에 따라 중재 효과의 차이가 통계적으로 유의함을 의미한다. 즉 이는 0으로 코딩된 집단에 비해 1로 코딩된 집단이 중재효과가 유의하게 더 크거나 작음을 뜻한다. 학생유형 중에서는 유의수준 0.05에서 지적장애 유무에 따라 중재효과 크기가 유의하게 다른 것으로 나타났다(B = 0.532, p = 0.019). 즉 지적장애를 지닌 학습곤란 학생의 경우 효과크기가 1.304이었던 반면 지적장애를 지니지 않은 학습곤란 학생(학습장애 진단 및 위험, 저성취, 학습부진 등)의 경우 효과크기가 1.836로, 데이터기반교수의 중재효과가 0.532 표준편차만큼 더 높다고 볼 수 있다. 학교급, 학습장애(진단 및 위험) 여부, 저성취 및 학습부진 여부, 성별, 학교유형 변수의 효과는 통계적으로 유의하지 않았다. 즉 연구참여자의 학교급과 관계없이, 학습곤란 학생 집단 내에서 학습장애(진단 및 위험), 저성취, 학습부진 해당 여부에 관계없이, 성별에 관계없이, 일반학급 여부에 관계없이 기초학습능력에 대한 데이터기반교수의 효과가 유의하다고 해석할 수 있다.

<표 Ⅳ-11> 학생 관련 예측변인 분석결과

고정효과		회귀계수	표준오차	t	자유도	유의확률
학교급		0.043	0.161	0.267	79	0.790
학생 유형	학습장애 진단/위험	-0.237	0.201	-1.184	79	0.240
	저성취 및 학습부진	-0.034	0.177	-0.195	79	0.846
	지적장애	0.532	0.222	2.394*	79	0.019
성별		0.352	0.177	1.988	79	0.050
학교 유형		0.312	0.184	1.694	79	0.094

*p<.05 **p<.01 ***p<.001

둘째, 중재 관련 변인의 조절효과 분석 결과는 <표 Ⅳ-12>와 같다. 유의수준 0.05에서 전체 회기 수(B = -0.568, p = 0.004)와 1회기 당 중재 시간(B = 0.608 p = 0.005) 변인이 통계적으로 유의했고, 나머지 변인에 대해서는 중재 효과 크기에 유의한 차이가 없는 것으로 나타났다. 전체 회기 수의 경우, 데이터기반교수가 19회기 이하로 제공되었을 때 효과크기가 1.552이었던 반면, 더 오랜 기간 동안 중재가 이어진 경우에는 효과크기가 0.984로 오히려 0.568 표준편차만큼 더 낮아지는 것으로 나타났다. 또한 1회기 당 중재시간을 45분 기준으로 구분하여 보았을 때 45분 이하인 경우에는 효과크기가 1.281이었고, 45분 초과인 경우에는 효과크기가 1.889인 것으로 나타나, 회기 당 시간이 더 길게 제공되었던 참여자에게 중재 효과가 0.608 표준편차만큼 커지는 경향이 있었다. 다만 전체 회기 수와 1회기 당 중재시간은 더미코딩으로 분류된 집단 간 효과크기 개수 차이가 큰 편이었으므로 신중한 해석이 필요하다. 나머지 변수에 대해서는 유의한 차이가 없었는데, 다시 말해 주당 회기수(1회기 또는 2~3회기), 중재 기간(9주 이하 또는 초과), 집단 크기(일대일 또는 2

명 이상)와 관계없이 데이터기반교수의 효과가 유의했다고 할 수 있다.

<표 Ⅳ-12> 중재 관련 예측변인 분석결과

고정효과	회귀계수	표준오차	t	자유도	유의확률
전체 회기 수	-0.568	0.189	-3.001**	79	0.004
주당 회기 수	-0.222	0.225	-0.989	79	0.325
1회기 당 중재 시간	0.608	0.212	2.870**	79	0.005
중재 기간	-0.193	0.189	-1.019	79	0.311
집단 크기	0.006	0.217	0.026	79	0.979

*p<.05 **p<.01 ***p<.001

마지막으로, 데이터기반교수 관련 변인을 중재효과에 대한 모형에 투입한 결과는 <표 Ⅳ-13>과 같다. 분석 결과, 진전도 측정 주기, 데이터기반 의사결정 유형, 교수적 수정 내용의 제시 관련 변인 모두 통계적으로 유의하지 않은 것으로 나타났다. 즉 진전도를 매회기 측정한 경우와 2회기 이상마다 측정한 경우를 비교했을 때 중재 효과의 유의한 차이가 없었다. 또 데이터기반 의사결정 유형과 관련하여 학생의 발달선과 목표선의 기울기를 비교하여 의사결정을 내린 경우와 기울기를 고려하지 않고 숙달기준, 백분위점수, 또래규준에 기초하여 의사결정을 내렸거나, 또는 관련 내용을 제시하지 않은 경우로 구분하여 비교하였을 때도 유의한 차이 없이 모두 효과적이었다. 마지막으로 학생의 데이터를 바탕으로 어떤 교수적 수정을 가했는지 내용을 명시적으로 제시한 경우와 제시하지 않은 경우의 비교에서도 중재 효과의 유의한 차이는 나타나지 않았다.

<표 Ⅳ-13> 데이터기반교수 관련 예측변인 분석결과

고정효과	회귀계수	표준오차	t	자유도	유의확률
진전도 측정 주기	-0.164	0.185	-0.886	79	0.378

데이터기반 의사결정 유형	-0.356	0.207	-1.714	79	0.090
교수적 수정 내용 제시	-0.301	0.194	-1.547	79	0.126

*p<.05 **p<.01 ***p<.001

Ⅴ. 논의 및 결론

1. 논의 및 결론

본 연구에서는 다양한 이유로 기초학습에 심한 어려움을 보이는 학령기 학습곤란 학생들을 대상으로 효과적이라고 제시되는 데이터기반교수의 효과를 확인하고 질적 수준을 파악함으로써 증거기반실제를 위한 자료를 제공하고자 하였다. 이를 위해 32편의 단일대상설계 연구를 중심으로 질적 분석 및 효과크기 분석을 실시하였다. 본 연구의 결과를 바탕으로 도출된 논의는 다음과 같다.

첫째, 2000년대부터 시작하여 국내에서 학령기 학습곤란 학생을 대상으로 데이터기반교수를 실시한 연구들이 지속적으로 이루어져 옴을 확인할 수 있었다. 이는 다양한 요인으로 학습에 어려움을 겪는 학습곤란 학생들에 대한 인식과 지원에 대한 필요성을 바탕으로 데이터기반교수를 현장에 적용하려는 노력이 지속적으로 있었음을 시사한다. 특히 연구의 설계와 관련하여, 대부분의 연구들이 단일대상설계 연구의 형태로 수행되어 왔음을 확인하였는데, 이는 특수교육계 연구 상황과도 관련이 있다. 즉 현실적으로 학습에 심각한 어려움을 겪는 학습곤란 학생 집단을 많은 수로 구하기 어려울 뿐 아니라, 집단 내에서도 개별적인 특성 차이가 커 동일한 중재를 제공하기 어렵고(김애화 외, 2011), 또 데이터기반교수의

핵심적 원리 중 하나가 지속적이고 빈번한 성취도 측정임을 고려할 때 단일대상설계와 잘 부합한다고 볼 수 있다. 단일대상설계 연구에서는 수준, 경향, 변동성, 효과의 즉시성, 중복, 인접한 단계에 걸친 자료의 변동성 등 시각적 분석을 통하여 독립변수와 종속변수 간 관계를 설명하는데(Kennedy, 2005; Horner et al., 2012), 기울기 또는 데이터 포인트 개수로 학생의 진전을 지속적으로 점검하는 데이터기반교수는 이러한 시각적 분석을 용이하게 할 것이다.

한편 분석대상논문들의 주요 특징을 살펴보면, 대부분의 연구가 초등학교 학생을 대상으로 일반학교에서 실시됨을 알 수 있었다. 목표 기술로는 읽기유창성이 가장 많았는데, 이는 실제로 80% 이상의 학습장애 학생들이 읽기 영역에서 어려움을 보고하고 있으며(Kim et al., 2017), 특히 읽기유창성은 여러 언어적 능력이 통합되어 있어 읽기 능력을 예측하는 정확한 지표로 평가된다는 점과 연관이 있을 것이다(Abbott et al., 2012; Fuchs et al., 2001; Katzir et al., 2006). 중재 관련 특성으로는 총 10~19회기, 주 2회, 1회기 당 31~45분으로 이루어지는 경우가 가장 많았다. 데이터기반교수와 관련하여서는 읽기유창성 CBM이 진전도 점검 도구로 가장 빈번하게 활용되었고, 진전도를 매회기 측정하는 경우가 가장 많았다. 데이터기반 의사결정 및 교수적 수정과 관련하여서는 정보의 누락이 많았는데, 내용을 기술한 연구 중에는 기울기 방법과 데이터포인트 방법을 함께 사용하고 질적인 측면에서 교수적 수정을 한 경우가 많았다.

그러나 이와 같이 데이터기반교수를 현장 및 연구에서 적용하려는 노력들이 이루어져왔음에도 불구하고, 아직까지 많은 연구들에서 데이터기반교수의 절차에 대한 구체적인 정보들을 누락하고 있었다. 구체적으로는 분석대상논문 중 약 절반에 달하는 연구에서 교수적 수정과 데이터기반 의사결정에 대한 구체적 정보를 제시하지 않음을 확인할 수 있었다. 증거에 기반한 교수적 수정은 특수교육의 주요한 특징인 개별화에 있어 현재까지 가장 검증된 접근법이자(신재현, 2019c), 교수적 수정의 양과 질은 학습곤란 학생의 성취에 유의한 영향을 준다고 보고된다(Stecker,

Fuchs, & Fuchs, 2005). 또한 교수적 수정은 학습곤란 학생의 성취뿐 아니라, 교사에게도 중요한 역할을 한다. 실제로 학생의 요구에 완벽하게 부합하는 중재 프로그램은 있을 수 없기에 교사가 어떤 프로그램을 선택한 이후에도 자신의 교수 전략을 평가하고 수정을 하는 역량이 강조되기 때문이다(신재현, 2019c). 실제로 McMaster 외(2020)의 연구에서 데이터기반교수 원리를 따른 교사들은 그렇지 않은 교사들에 비해 더 빈번하게, 또 교사의 직관이 아닌 객관적인 데이터에 근거하여 교수적 수정을 하였으며, 환경, 전달방식, 자료, 초점 등 보다 다양한 측면에서 교수적 수정을 한 것으로 나타났다. 데이터기반 의사결정의 경우, 선행연구에 따라 기울기 방법과 데이터 포인트 방법, 또는 이에 더하여 숙달기준 방법이 제안된다(Ardoin et al., 2013). 그러나 분석대상논문 중 많은 수에서 학생의 성취 데이터를 바탕으로 목표선과 추세선을 그려 교수 계획에 적용하였다는 점은 긍정적이나, 어떠한 구체적인 준거에 의해 교수적 의사결정을 내렸는지를 제시하지 않은 연구가 많았다. 또 학생의 수행수준이 기준보다 낮거나 높게 나타나도 교수적 수정을 하지 않고 기존의 중재를 그대로 지속하는 경우도 많았다. 이러한 경우에는 비록 중재의 결과로 학생의 기초학습능력 수준이 유의미하게 향상되어 중재 효과가 입증되었다고 하더라도, 어떠한 구체적인 데이터기반교수 절차에 의해 효과가 있었는지 확인하고 추후 연구를 반복하는 데 제한이 있다. 따라서 추후에는 데이터기반교수를 적용하여 연구를 할 때 어떠한 구체적인 기준과 절차에 의해 중재를 수정하고 교육적 의사결정을 내렸는지 더욱 충분히 제시할 필요가 있을 것이다.

둘째, 분석대상논문의 질적 지표 분석 결과, Horner 외(2005)의 22개 지표 중 14개에서 전체 연구에 걸친 평균 점수가 2점 이상으로 분석되었고, 전체 분석대상논문 중 2편을 제외한 30편의 연구에서 질적 지표의 평균 점수가 2점 이상으로 분석되었다. 이를 통해 현재까지 국내에서 학령기 학습곤란 학생을 대상으로 데이터기반교수를 실시한 연구가 충분한 질적 수준을 갖춘 것으로 판단할 수 있었다. 증거기반실제를 확인하기 위해서는 일반적으로 단일대상설계 연구의 경우 5편 이상 또는 연구참여

가자 최소 20명 이상이 되어야 한다(Horner et al., 2005). 따라서 본 연구의 결과는 국내 특수교육 및 일반교육에서 데이터기반교수가 증거기반 실제로 고려될 수 있으며 적극적으로 활용되어야 함을 시사한다.

구체적으로 질적 지표 중 연구참여자 기술과 관련하여 거의 모든 연구에서 연구참여자에 대한 정보, 선별 과정 및 기준, 학습곤란의 특성 등을 자세히 보고하고 있었다. 또한 종속변수 기술 및 측정과 관련하여서도 대부분의 연구에서 높은 질적 수준을 충족하고 있는 것으로 보고되었는데, 이는 데이터기반교수에서 그 중요성이 강조되는 신뢰로운 검사도구를 통해 양적 자료를 수집하고 CBM 원리에 의해 높은 빈도로 반복 측정을 실시한 것과 관련이 있다.

그러나 더 나은 질적 수준의 확보가 필요한 측면도 있는 것으로 분석되었다. 질적 지표 분석에서 평균 점수가 2점 이하로 불충분하게 보고된 지표로는 중재제공자 기술, 중재 장소 기술, 종속변인의 측정 신뢰도 및 관찰자간 일치도 측정 및 기술, 연구자 통제 하 독립변수 조작, 중재 충실도 측정 및 기술, 기초선 기간 기술, 독립적인 실행의 실용성 및 비용효과성 기술, 독립적인 수행 특성 기술이 있었다. 특히 중재제공자 및 중재 장소와 관련하여, 많은 연구에서 중재자가 누구인지(연구자, 학급 담임 교사, 부진아 지도 교사 등), 중재 장소가 어디인지(학생이 속해 있는 통합학급 또는 특수학급, 방과후 교실, 특수학교 특별실 등)와 같은 형태로 간략하게 제시하는 데 그쳤다. 중재제공자의 경력, 성별, 교육수준, 학습곤란 학생 중재에 대한 친숙도 등의 구체적 정보와 중재 장소의 책상 배치, 크기, 학생과 교사의 수 등의 자세한 정보는 중재의 외적 타당도에 필수적인 반복(replication)을 위해 중요하므로(서유진, 나경은, 2012), 이에 대한 기술이 더욱 충분히 이루어질 필요가 있다. 또 측정신뢰도, 중재 충실도는 연구결과의 신뢰성을 위해 필수적인 요인이며(Horner et al., 2005), 중재 계획 수립 및 결과 평가에 있어 교사, 부모 등 연구자 이외의 개인이 참여하여 중재의 필요성 및 적용가능성 여부에 대한 평가를 하는 사회적 타당도 역시 연구의 질을 결정하는 중요한 요인이다(김은경, 김라경, 2014; 양명희, 김황용, 2002). 이외에도 연구자의 체계적인 실

험 통제는 단일대상설계에서 독립변수의 조작과 종속변수의 변화 간 기능적 관계를 입증하기 위해 강조된다(Horner et al., 2005). 따라서 앞으로의 데이터기반교수 연구에서는 이와 같은 사항들을 충실히 제시함으로써 연구 결과의 신뢰도를 높이고 데이터기반교수의 효과를 보다 객관적으로 입증할 필요가 있겠다.

셋째, 데이터기반교수의 효과를 단일한 효과크기로 종합하기 위하여 1수준을 시점 측정치, 2수준을 연구참여자로 설정하여 다층메타분석을 실시한 결과, 학령기 학습곤란 학생 대상 데이터기반교수의 효과는 상당히 높은 수준으로 나타났다. 구체적으로 데이터기반교수의 표준화된 효과크기는 1.34로, 통계적으로 유의했다. 이는 학령기 학습곤란 학생을 대상으로 데이터기반교수를 실시한 결과, 기초선 구간에 비해 중재 구간에서 기초학습능력 수준이 평균적으로 1.34 표준편차만큼 향상되었음을 의미한다. 또한 기초학습능력 영역을 읽기, 쓰기, 수학으로 구분하여 확인한 결과, 각각 1.43, 1.52, 1.09로 높은 수준이었고 모두 통계적으로 유의하였다. 6편의 (준)집단실험설계 연구에서도 Hedge's *g*로 표현한 효과크기의 평균은 1.54, 중앙값은 1.20으로 나타나, 전체적으로 매우 높은 효과크기에 해당되었다. 비록 단일대상설계 연구의 효과크기는 집단설계 연구 메타분석에서의 효과크기와 직접적으로 비교가 불가능하지만, 본 연구 결과를 통해 데이터기반교수가 학습곤란 학생들의 성취를 향상하는 데 효과가 높음을 확인할 수 있다.

일반적으로 학습곤란 학생들은 학습 기술을 습득하고 발전시키는 속도가 또래보다 느린 편이다(Bast & Reitsma, 1998; Francis, 1992). 특히 심각한 학습곤란을 지닌 학생들의 경우, 성취를 높이는 데 중재 빈도 및 시간을 늘리거나 교수집단의 크기를 줄이는 등 일반적인 중재 강화 방식은 충분치 않을 수 있다고 보고된다(Fuchs & Fuchs, 2015; NCII, 2013). 그럼에도 불구하고 본 연구에서 표준화된 효과크기가 1.34의 높은 수준으로 나타난 것은 이들에게 데이터기반교수의 원리에 따라 중재를 충실히 제공하고 적절한 도구로 중재에 대한 반응을 측정할 때 충분한 진전을 보일 수 있음을 시사한다. 학습곤란 학생에게 있어 데이터기반교수의

높은 효과는 몇 가지 측면에서 이해될 수 있는데, 우선 학생 개개인에 대한 개별적 접근이 가능한 것과 더불어, 교사의 임의적인 판단이 아닌 객관적인 데이터를 바탕으로 타당한 의사결정을 내리는 것과 연관이 있을 것이다. 또한 CBM 도구의 활용은 학습곤란을 지닌 학생의 능력을 측정하는 데 매우 적절하다고 여겨지는데(Shin et al., 2004), 이를 통해 학습곤란 학생의 성취 진전이 더욱 민감하고 정확하게 측정되었다고 볼 수 있다. 전술한 바와 같이 이들이 상대적으로 느리고 작은 폭으로 성장한다는 점을 고려하면, 일반적인 학업성취검사로는 중재의 효과나 진전을 민감하게 측정하기 어렵다. 예컨대 읽기 능력을 측정한다고 하였을 때 실제 속한 학년의 교과에서 다루는 지문을 소리 내어 읽거나 이해하도록 하는 것은 이들의 현재 능력 수준 밖에 있으며, 중재 결과 이들의 성취가 실제로 향상되었음에도 불구하고 적절한 측정도구의 부재로 인해 그것을 민감하게 식별하지 못할 수도 있다. 반면 데이터기반교수에서 사용되는 CBM 읽기검사의 경우 여러 학년에 걸쳐 동일한, 쉬운 수준의 텍스트로 읽기능력을 측정하여, 마치 아동기 동안 키와 체중의 성장을 지속적으로 확인하듯이 개별 학생이 성취 진전을 확인할 수 있다(Deno et al., 2001). 특히 학습장애 선별 및 진단 모형인 중재반응접근에서 '효과적인 중재에도 반응하지 않음'을 확인하고자 한다면, 데이터기반교수가 선행되었으며 반응도를 적절한 도구로 측정하였는지에 대한 확인이 필요할 것이다. 또한 특수교육대상자로 진단받지 않거나 못하여 일반교육과 특수교육의 사각지대에 놓인 학습곤란 학생들이 많음을 고려할 때, 일반학급에서부터 일반교육과 특수교육이 협력하여 효과적인 데이터기반교수를 제공할 필요가 있을 것이다.

다만 본 연구에서 산출된 효과크기는 기존에 집단실험설계 연구를 중심으로 증거기반개별화교수의 효과를 확인한 Jung 외(2018)의 결과(Hedge's g = 0.37 ~ 0.38) 및 증거기반 의사결정을 적용한 읽기 중재의 효과를 확인한 Filderman 외(2018)의 결과(Hedge's g = 0.24 ~ 0.27)에 비해 상당히 높은 수준으로 해석될 수 있다. 다만, 먼저 본 연구에서 계산된 단일대상설계 연구의 표준화된 효과크기는 집단설계의 경우와 산

출식이 달라 서로 직접적으로 비교하기는 어렵다는 점에 주의해야 한다(Wang, Cui, & Parrila, 2011). 또 이러한 효과크기의 차이는 몇 가지 연구 간 차이와 관련이 있다. 예컨대 분석대상에 포함된 연구들의 특성으로 인해 차이가 발생하였을 수도 있는데, 구체적으로는 Jung 외(2018)에서는 심각한 학습 요구(intensive learning needs)를 지닌 학생에 초점을 두기 위해 학습장애, 지적장애, 정서행동장애 등 장애로 진단받은 학생을 대상으로 한 연구만을 종합하였고, Filderman 외(2018)에서는 학령기 학생을 대상으로 한 본 연구와 달리 유치원생을 포함하여 읽기 영역에 한정된 중재 효과를 확인하였다는 점에서 다르다. 또한 국외 메타분석 연구의 경우 중재 및 모니터링이 7주 이상 동안 이루어졌을 것(Jung et al., 2018), 중재 빈도 및 간격, 집단 크기, 교수자료의 개별화, 교수적 의사결정의 내용 및 그를 위한 데이터 활용 절차가 제시되어 있을 것을 전제로 하는 등(Filderman et al., 2018) 문헌 선정 기준을 상대적으로 엄격하게 설정하였다. 이러한 몇 가지 차이는 산출된 효과크기의 차이에 영향을 미친 것으로 보인다.

넷째, 단일대상설계 연구의 분석 결과, 학령기 학습곤란 학생들 간 데이터기반교수의 효과는 동질적이지 않은 것으로 나타났다. 연구참여자 간 효과크기의 이질성을 설명하는 변인이 무엇인지 탐색하고자 2수준에 예측변인을 투입한 결과, 학생유형, 전체 회기수, 1회기 당 중재 시간에 따른 유의한 차이가 있었다. 구체적으로 학생유형에서는 학생의 지적장애 유무에 따른 효과크기 차이가 있었는데, 지적장애가 있는 학습곤란 학생에 비해 지적장애가 없는 학습곤란 학생(학습장애 진단 및 위험, 저성취, 학습부진 등)의 경우 데이터기반교수의 중재효과가 0.532 표준편차만큼 더 높은 것으로 나타났다. 그 외에는 학습곤란을 지닌 학생 집단 내에서 학습장애(진단 또는 위험)와 저성취 및 학습부진의 유무에 따른 효과 차이는 유의하지 않은 것으로 나타났다. 또한 학생의 학교급(초등 또는 중고등), 성별(남자 또는 여자), 학교유형(일반학급 또는 일반학급 외)에 따른 차이도 유의하지 않았는데, 이는 학습곤란 학생의 학교급, 성별, 학교유형에 관계없이 데이터기반교수가 기초학습능력 증진에 미치는

효과가 유의하다고 이해할 수 있다. 이는 학생 집단에 따라 CBM 기울기에 차이가 있는지를 검증한 선행연구들과도 맥을 같이 하는데, Chard 외(2008), 여승수, 정광조(2019)의 연구에서는 성별, 인종, SES 등 학생변인이 CBM 기울기에 유의한 영향을 주지 않는 것을 확인하였다. 이러한 결과는 학교 현장에 존재하는 다양한 특성과 교육적 요구를 지니는 학습곤란 학생들을 위하여 데이터기반교수를 적극적으로 활용할 수 있음을 시사한다.

한편 지적장애 유무에 따른 데이터기반교수 효과의 차이와 관련하여, 지적장애를 지닌 학생들의 인지적 특성이 상대적으로 낮은 효과크기에 영향을 미친 것으로 보인다. 본 연구의 분석대상논문 중 지적장애를 지닌 연구참여자의 수는 14명이었고 그에 따른 효과크기의 수는 총 17개였다. 이들의 지능지수 평균은 55.9로 경도 지적장애 범위에 해당되는 것으로 나타났다. 선행 연구에 따르면 경도 지적장애 학생들은 다른 학생들과 구분되는 인지적 특성 및 그에 따른 교육적 요구를 지니는데(Sabornie et al., 2005), 주의집중 및 유지, 일반화 및 전이, 정보 입력 및 인출 등에서 어려움을 나타낸다고 보고된다(Kirk et al., 2011). 따라서 이러한 특성이 비교적 낮은 진전에 영향을 미쳤다고 볼 수 있다. 다만 본 연구에서 분석에 활용된 지적장애를 지닌 연구참여자 및 효과크기의 수가 적은 편이므로 신중한 해석이 요구된다.

오히려 지적장애를 지닌 학생 집단에서도 데이터기반교수의 효과크기는 유의한 것으로 나타났고, 1.304로 상당히 높은 수준이었다는 점에 주목할 필요가 있다. 이러한 결과는 지적장애가 있는 학생의 학습능력 증진을 위하여 다양한 중재 전략들이 연구, 적용되고 있는데, 그 중 하나로 데이터기반교수의 활용 가능성을 시사한다. 데이터기반교수는 통합학급교사 또는 특수교사로 하여금 지적장애 학생을 위해 구체적이고 현실적인 교수 목표를 설정하고, 손쉽게 신뢰로운 데이터를 수집하며 지속적인 모니터링을 할 수 있도록 돕는다. 특히 지적장애 학생의 학습능력 측정에 있어 CBM의 적용 가능성에 주목할 필요가 있다. 지적장애 학생의 기초학습능력 측정을 위해 다른 학생에게 사용하는 것과 동일한 평가를

적용하는 것은 이들로 하여금 자신이 아는 것을 보여주는 데 제한이 될 수 있다. Snyder & Ayres(2020)에 따르면 미국에서 읽기 능력을 측정하기 위해 주로 사용되는 Woodcock Reading Mastery Test(WRMT; Woodcock, 2011) 및 Comprehensive Test of Phonological Processing(CTOPP-2; Wagner et al., 2013) 등은 지적장애를 가진 학생에게 난이도가 지나치게 높을 뿐 아니라, 많은 지적장애 학생들의 경우 이러한 유형의 평가 항목에 대한 경험이 제한적이므로 적절하지 않다. 이에 비해 CBM 검사와 같은 일반적인 성취도 측정 도구(general outcome measurement)의 경우 지적장애 학생의 읽기 수준과 진전을 확인하는 데 있어 보다 효과적이라고 제시된다(Wallace, Ticha, & Gustafson, 2008). 또한 CBM은 읽기, 쓰기, 수학 영역에서 지적장애 학생의 개별화교육계획 목표를 작성하는 수단으로서도 활용될 수 있다고 제시되는데, CBM을 바탕으로 한 교육목표는 작성하기 쉽고, 이해하기 쉬우며, 진행상황을 모니터링하기 유용하기 때문이다(Mattatall, 2011). 구체적으로 지적장애를 지닌 학생을 위한 CBM의 실제적 활용 방안과 관련하여, Lemons 외(2013)에서는 CBM 검사의 가장 낮은 학년 단계에서 시작하여 지적장애 학생의 성취가 규준집단의 50번째 백분위점수와 같이 미리 설정한 일정 기준보다 높은 것으로 나타날 때 다음 학년 단계 검사로 넘어가는 방법을 사용한 바 있다. 이를 종합하여 볼 때 경도 수준의 지적장애를 지닌 학습곤란 학생을 대상으로도 데이터기반교수는 효과적이며, 적절한 목표 설정, 측정 도구의 적절성, IEP 작성에의 활용 등의 측면에서 활용 가능성이 높다. 특히 일반 교육과정으로의 통합이 더욱 강조되는 경도 수준의 지적장애학생에게 데이터기반교수가 적용될 수 있을 것이다.

다음으로 전체 회기 수 및 1회기 당 중재 시간에 따른 효과 차이와 관련하여 살펴보면 다음과 같다. 본 연구에서는 19회기 이하로 제공되었을 때에 비해 더 오랜 기간 동안 중재가 이어진 경우에는 효과크기가 오히려 0.568 표준편차만큼 더 낮아지는 것으로 나타났다. 또한 1회기 당 중재시간을 45분 기준으로 구분하여 보았을 때, 회기 당 시간이 더 길게

제공되었던 참여자에게 중재 효과가 0.608 표준편차만큼 커지는 경향이 있었다. 중재 기간이 길어질수록 효과크기가 낮게 나온 것은 데이터기반교수의 기간이 학생의 쓰기 능력 향상을 정적으로 예측하는 변인으로 나타난 Bresina & McMaster(2020)의 연구와는 상반된 결과이다. 회기 당 중재 시간과 관련하여서는, 학생이 교사로부터 얼마나 많은 피드백, 오류 교정을 받는지는 회기 당 교수 시간과 밀접한 관련이 있고(신재현, 2019), 회기 당 중재 시간이 길어질수록 피드백 횟수가 늘어 중재효과의 크기가 커진 것으로 예상할 수 있다. 다만 데이터기반교수에 기초한 쓰기 중재를 실시한 McMaster 외(2020)에서는 교사들로 하여금 최소 주 3회, 한 회기 당 20~30분의 비교적 짧은 시간 동안 중재를 실시할 것을 권유한 바 있다. 아직까지 효과적인 중재 기간, 회기 당 시간, CBM의 측정 주기 등에 대해서는 지속적인 연구가 필요한 실정이다(여승수, 정광조, 2019; Jung et al., 2018). 더불어 학생의 개별 특성, 사용된 CBM 측정도구의 특성에 따라서도 효과크기에 영향을 미칠 수 있으므로 연구 결과 비교 시 세심한 고려가 필요할 것이다.

다섯째, 데이터기반교수 실시에 있어 교사에 대한 지원은 데이터기반교수의 효과를 유의하게 높이는 변인이고, 교사 지원의 제공 여부, 빈도, 형태 등에 따라 효과의 유의한 차이가 있다고 보고된다(Jung et al., 2018; McMaster et al., 2020). 이에 국외의 경우 데이터기반교수를 위한 코칭(coaching), 개인 및 소집단 컨설팅 등 교사 지원을 위해 추가적인 시간이 할애되고 데이터기반교수의 충실도를 위해 체계적인 과정을 거친다. 그러나 본 연구의 분석대상논문 중 교사 지원이 이루어진 경우는 거의 없는 것으로 나타났다. 이에 연구에서 제시하고 있는 데이터기반교수의 효과적인 원리 및 절차가 교육 현장에서 잘 적용될 수 있도록 교사를 대상으로 데이터기반교수에 대한 교육, 중재 기간 동안의 주기적인 점검 및 지원, 관련 자료 제공 등 다양한 지원이 적극적으로 요구된다. 실제로 현장에 있는 교사들은 개별 학생의 특수한 교육적 요구에 맞추어 교수를 제공하고자 헌신한다. 그럼에도 불구하고 많은 경우 교수 수정에 있어 교수 수정에 대한 지식 및 정보의 부족으로 인한 어려움을 호소하는 것

으로 나타난다(최세민, 권택환, 2009). 따라서 증거가 충분한 효과적인 교수전략으로는 무엇이 제시되고 있는지, 어떤 기준에 근거하여 교수적 수정을 할 수 있는지 등 데이터기반교수에 대한 교사의 이해를 높이기 위한 지원이 필요하며, 이러한 데이터기반교수에 대한 전문적 이해는 비단 특수교사 뿐 아니라 일반교사에게도 강조되어야 할 것이다. 특히 학교 현장의 교사들은 학생들 가까이에서 이들과 직접적으로 상호작용하며 학습곤란을 겪는 학습자를 1차적으로 발견하는 중요한 주체임을 고려할 때(최수미 외, 2018), 이러한 교사 지원의 중요성은 매우 크다.

더불어 교사들이 데이터기반교수와 관련된 지식 및 정보에 보다 손쉽게 접근할 수 있는 창구가 필요할 것이다. 이대식(2019)는 국내 학습장애 분야의 위기와 관련하여, 그 가장 큰 본질은 진단 및 판별 상의 어려움도, 일반교육과 특수교육 간 분리도, 장애 진단에 대한 기피도 아닌, 교사들에게 실제 학습장애 학생과 마주 보고 앉아 지도하기 위한 증거기반 교수방법, 도구, 자료를 충분하고 접근 가능한 형태로 제시해주지 못했기 때문이라고 보았다. 이는 비단 학습장애 뿐 아니라 다양한 학습곤란 학생들을 지도해야 하는 교사들에게 그들이 체감할 수 있는 정도의 연구에 기초한 실천적 대안이 제공될 필요가 있음을 시사한다. 미국의 경우 이러한 시도가 보다 적극적으로 이루어져 왔다. 예컨대 미국 교육부의 'What Works Clearninghouse(WWC) 플랫폼에서는 기초학습영역별 데이터기반교수를 검색, 비교, 평가할 수 있도록 하고 전문적 이해 없이도 직관적으로 이해하기 쉬운 정보를 제공하고 있으며, 미국 교육부 산하 National Center on Intensive Intervention(NCII)에서는 교사를 대상으로 진전도 점검을 위한 도구 차트(tool chart) 및 교수 수정 유형별 체크리스트 등을 제공하고 있다(신재현, 2019c). 이와 같이 연구에 기초한 데이터기반교수가 현장에 잘 전달될 수 있도록 전달의 주체가 되는 교사를 위한 지원이 필요하고, 이는 연구와 현장을 연결하고 데이터기반교수의 효과를 높이는 데 중요한 역할을 할 것이다.

2. 연구의 제한점 및 제언

본 연구의 제한점 및 후속 연구를 위한 제언은 다음과 같다.

첫째, 본 연구에서는 데이터기반교수의 핵심 원리와 절차를 반영한 문헌을 분석대상으로 선정했음에도 불구하고, 상당수의 연구에서 교수적 수정 내용, 데이터기반 의사결정의 기준 및 구체적 절차 등을 보고하지 않고 있어 분석에 제한이 있었다. 특히 본 연구에서는 데이터기반교수의 효과에 차이를 가져오는 변인으로 학생유형, 전체 회기수, 1회기 당 중재시간이 확인되었으나, 추후 데이터기반교수 관련 변인들에 대해 더욱 충분한 정보가 확보되었을 때 여러 예측변인의 조절효과를 확인, 비교할 필요가 있겠다. 실제로 집단설계연구를 중심으로 메타분석을 실시한 Jung 외(2018)의 연구에서는 CBM 측정 주기, 교사 지원의 유형 및 빈도, CBM 검사 종류(교사개발 CBM과 연구자 개발 CBM), 증거기반 의사결정 준거, 데이터기반 의사결정에 활용된 데이터 수 등에 따라 효과크기에 차이가 있는 것으로 나타났다. 따라서 학령기 학습곤란 학생을 대상으로 데이터기반교수를 적용하고 보다 구체적인 정보를 포함한 연구들이 지속적으로 축적될 필요가 있고, 그 후 더욱 다양한 예측변인의 효과를 탐색할 수 있을 것으로 기대한다.

둘째, 본 연구에서는 단일대상설계 연구의 3수준 다층메타분석을 실시하기에는 연구 편수가 부족하여, 2수준 분석에 그쳤다. 추후 더욱 충분한 연구 수와 예측변인 관련 정보들이 확보된다면, 3수준 다층메타분석을 실시하여 연구참여자 간 효과크기의 이질성 뿐 아니라 연구 간 효과크기의 이질성을 확인하며, 개인 수준 특성과 연구 수준 특성을 구분하여 각각의 특성이 효과크기의 차이에 미치는 영향을 탐색할 수 있을 것이다. 또 (준)집단실험설계의 경우 6편에 불과하여 효과크기를 종합하지 않았는다. 앞으로는 국내에서 학령기 학습곤란 학생에게 데이터기반교수의 효과를 검증하는 연구가 집단설계 형태로도 활발하게 이뤄질 필요가 있고, 더불어 충분한 연구를 바탕으로 평균 효과크기를 도출하여 선행연구와 비교할 수 있기를 기대한다.

셋째, 본 연구에서는 Van den Noortgate & Onghena (2003a; 2003b; 2007)에서 제시하고 Wang, Cui, & Parrila (2011), Wang, Parrila, & Cui (2013), Heyvaert 외(2012) 등에서 제시한 방법으로 단일대상연구의 효과크기를 종합하였다. 그러나 아직 단일대상 메타분석의 효과를 종합하는 합의된 방법이 부재하므로 보다 다양한 분석 방법으로 효과크기를 산출하여 비교해볼 필요가 있다. 예컨대 Swanson & Sachse-Lee (2000)에서는 두 개 이상의 단일대상설계 연구들을 종합하여 표준화된 효과크기를 구하기 위해 기초선과 중재 단계의 마지막 3회 측정치를 비교하였는데, 이는 마지막 3회기에서는 어느 정도의 안정화가 이루어진다고 보았기 때문이다. 또 두 집단의 차이를 나타내는 효과크기에 비해 중재 전후 비교의 경우 효과크기가 더 증대될 수 있기 때문에 사전, 사후 점수의 상관관계를 활용하여 효과크기를 보정하는 단계를 거쳤다. 이와 같이 보다 다양한 방법으로 도출된 결과를 비교하고 각 분석방법의 강약점을 고려하여 해석할 필요가 있을 것이다. 또한 단일대상설계 연구의 경우 연구참여자의 특성, 개별화된 중재의 특성 등 개별 연구마다의 특성이 매우 다양하다. 따라서 양적으로 종합된 효과크기를 해석하는 것을 넘어 개별 연구의 특징이 함께 고려될 필요가 있을 것이다.

넷째, 본 연구에서는 BASA 검사를 적용하여 모니터링을 실시한 논문만을 분석에 포함하였는데, 이는 국내에서 기초학습영역에서 CBM 원리에 기초하여 개발된 검사 중 해당 검사가 상용화된 표준화 검사도구이자 가장 빈번하게 활용되기 때문이다. 그러나 앞으로는 기초학습영역에서 학습곤란을 지닌 학생의 성장을 민감하게 확인하면서도 그 측정학적 적합성이 검증된 CBM 도구가 더욱 다양하게 개발되고 표준화 과정을 거칠 필요가 있을 것이다. 이를 바탕으로 추후 연구에서는 BASA 외에도 다양한 CBM 측정도구를 사용하여 데이터기반교수를 실시한 논문을 분석에 포함해야 할 것이다. 특히 본 연구에서는 연구자 및 교사 개발에 의한 CBM 도구의 측정학적 적합성이 낮다고 보고된다는 점을 고려하여, 보다 신뢰로운 결과를 위해 연구자 및 교사 개발에 의한 CBM 도구만으로 모니터링을 한 경우에는 분석대상논문에서 제외하였다. 그러나

추후 연구자 및 교사 개발 CBM 측정 도구 중에서도 그 타당도와 신뢰도가 비교적 높고 CBM 원리를 충실히 반영한 도구가 무엇이 있는지 확인할 필요가 있겠고, 그러한 검사도구들을 활용하여 데이터기반교수를 실시한 연구도 분석에 포함한다면 데이터기반교수의 효과를 보다 종합적으로 분석할 수 있을 것으로 보인다.

다섯째, 본 연구에서는 읽기, 쓰기, 수학 등 기초학습영역별 데이터기반교수의 평균 효과크기를 산출하였으나, 구체적인 교수 프로그램 및 교수전략에 따른 효과 차이를 확인하지는 못하였다. 데이터기반교수에서는 직접교수와 같이 증거가 확보된 교수 프로그램 및 교수전략을 사용하도록 권고된다. 따라서 추후 연구에서는 포괄적인 중재 영역을 넘어 구체적인 중재 내용 및 교수 전략에 따른 효과크기 차이를 확인함으로써 어떤 중재를 선택해야 하는지에 대한 정보를 보다 풍부하게 제공할 필요가 있을 것이다.

여섯째, 본 연구에서는 Horner 외(2005)의 지표를 활용하여 질적 분석을 실시하였으나, 후속 연구에서는 보다 다양한 기준을 바탕으로 데이터기반교수의 질적 분석이 이루어질 필요가 있다. 특히 신재현(2019c)은 국내외에서 특정 교수 프로그램을 적용한 연구의 질적 수준을 평가하고자 하는 시도가 많이 있어왔으나, 대부분 연구자의 입장에서 연구방법 측면의 충실도를 확인하는 한계가 존재한다고 지적한 바 있다. 따라서 본 연구에서 사용된 질적 지표는 연구들의 증거의 수준을 밝히는 데 중요한 역할을 할 수 있으나, 연구설계 및 방법론적 측면 외에도 데이터기반교수의 직접적 주체가 되는 교사 등 교육현장 전문가의 관점에서 데이터기반교수의 질적 수준을 평가할 필요가 있을 것이다. 예컨대, 미국에서 학교, 센터 등 교육 현장에서 교사들로 하여금 쉽고 적절한 데이터기반교수 선택을 지원하는 것을 목적으로 한 Fuchs 외(2017), National Center on Intensive Intervention(NCII) 등에서 제시한 데이터기반교수 평가기준을 참고할 수 있을 것이다. 추후 연구에서는 이와 같은 기준들을 함께 활용하여 연구설계 측면에서의 연구의 질 검증에 더불어, 선행연구에서 제시되는 데이터기반교수의 요소 및 절차를 얼마나 충실히 따르고 있는

가를 확인하며, 현장에의 적용가능성을 평가할 수 있을 것이다. 또한 교사들이 데이터기반교수를 실시함에 있어 어떤 어려움을 경험하는지, 어떤 기준으로 교수적 수정 및 교육적 의사결정을 하고 있는지 등을 보다 심층적으로 살펴볼 필요가 있다. 이를 통해 국내 학습곤란 학생 대상 데이터기반교수 연구의 증거기반실제로서의 수준을 더욱 다양한 관점에서 확인할 수 있을 것으로 기대한다.

참 고 문 헌

강옥려 (2016). 경계선급 지능 아동의 교육: 과제와 해결방안. **한국초등교육, 27**(1), 361-378.

강은영, 김우리, 신재현 (2020). 다양한 학업적 어려움을 지닌 위험군 학생을 위한 행정 및 제도적 지원체제 구축 방안과 과제. **특수교육학연구, 55**(3), 1-31.

강종구, 김라경 (2012). 사회구성주의 관점에서 본 지적장애의 개념과 시스템 고찰. **지적장애연구, 4**(2), 1-23.

고서연, 백영선, 길한아, 김미령, 손승현 (2018). 쓰기에 어려움이 있는 학생을 위한 쓰기중재 단일대상연구에 대한 메타분석: 경도지적장애, 학습장애, 학습부진학생을 중심으로. **특수교육저널: 이론과 실천, 19**(2), 123-148.

교육부 (2020). **2020 특수교육통계.** 세종: 교육부.

국립특수교육원 (2020). **2020 특수교육통계.** 아산: 국립특수교육원.

길한아, 손승현, 고서연, 백영선, 김미령, Sakuraa, Y. (2018). 읽기에 어려움이 있는 학생을 위한 읽기중재 단일대상연구에 대한 메타분석: 경도지적장애, 학습장애, 학습부진학생을 중심으로. **학습장애연구, 15**(3), 1-27.

김경미 (2013). 초등학교 읽기장애학생의 독해력 향상 연구에 관한 메타분석. 대구대학교 석사학위논문.

김동일 (1998). 교육과정중심평가의 이론과 실제: CBM(Curriculum-Based Measurement)을 중심으로. **현장특수교육, 5**(3), 8-17.

김동일 (2000). **기초학습기능 수행평가체제(BASA): 읽기.** 서울: 학지사.

김동일 (2002). 초등학생 쓰기 능력 진단과 지도를 위한 쓰기 평가와 쓰기 오류 분석. **아시아교육연구, 3**(1), 43-62.

김동일 (2006). **기초학습기능 수행평가체제(BASA): 수학.** 서울: 학지사.

김동일 (2009). **기초학습기능 수행평가체제(BASA): 쓰기.** 서울: 학지사.

김동일 (2011a). **기초학습기능 수행평가체제(BASA): 초기문해.** 서울: 학지사.

김동일 (2011b). **기초학습기능 수행평가체제(BASA): 초기수학.** 서울: 학지사.

김동일 (2018). **기초학습기능 수행평가체제(BASA): 수학 문장제.** 서울: 학지사.

김동일 (2019a). **기초학습기능 수행평가체제(BASA): 어휘.** 서울: 학지사.

김동일 (2019b). **기초학습기능 수행평가체제(BASA): 읽기이해.** 서울: 학지사.

김동일, 김미순, 배성직 (2003). 학습부진아동의 진단과 중재를 위한 쓰기 검사 체제 개발과 타당화. **아시아교육연구, 4**(3), 43-67.

김동일, 고혜정, 신재현, 김이내, 김붕년, 이기정 (2012). 특수교육지원센터의 학습장애 선별 및 진단과정에 대한 탐색적 연구: 현행 지침 시행과 개선방향에 대한 시사점을 중심으로. **아시아교육연구, 13**(3), 253-283.

김동일, 이재호, 이미지 (2013). 쓰기 학습장애 중재 단일대상연구에 대한 메타분석. **학습장애연구, 10**(2), 73-92.

김동일, 고혜정, 신재현, 김이내, 김우리야, 김붕년 (2013). 수학문장제 문제해결력 검사 타당화 연구: 학습장애 위험군 선별을 중심으로. **열린교육연구, 21**(1), 129-148.

김동일, 이대식, 신종호 (2016). **(DSM-5에 기반한) 학습장애아동의 이해와 교육 (3판).** 서울: 학지사.

김동일, 라수현, 이혜은 (2016). 메타인지전략의 효과에 관한 메타분석: 집단설계연구와 단일사례연구의 비교. **아시아교육연구, 17**(3), 21-48.

김동일, 안예지, 이미지, 조영희, 박소영, 고혜정 (2016). 기초학습 수행평

가체제 어휘검사 타당화 연구. **특수아동교육연구, 18**(3), 55-76.
김동일, 고혜정 (2018). 학습곤란(ld) 학생을 위한 교육지원의 확장적 전환: 학습장애(LD) 정의 체계 재구조화. **학습장애연구, 15**(1), 1-12.
김동일, 안예지, 김희은, 이연재 (2020). 중재반응모형(RTI) 기반 기초학력부진 학생 단계별 읽기중재 효과. **교육심리연구, 34**(2), 307-325.
김민경, 박유정 (2015). 학습부진 및 학습장애를 위한 어휘교수전략에 관한 고찰: 국내 연구 동향 및 효과 분석. **학습장애연구, 12**(2), 227-246.
김병룡 (2017). 초등 수학학습부진학생 및 학습장애학생을 대상으로 한 실험 및 준실험연구의 질적 수준 및 효과크기 분석. **초등교육연구, 30**(1), 47-72.
김우리, 고은영 (2012). 국내 학습장애 읽기이해 중재의 효과분석. **학습장애연구, 9**(3), 179-203.
김은경, 김라경 (2014). 국내 보완대체의사소통 (AAC) 중재가 특수교육대상학생의 의사소통 능력에 미치는 효과 분석. **특수교육, 13**(3), 25-51.
김은향, 김동일, 고은영 (2013). ADHD 위험군 아동의 쓰기 수행 수준과 오류유형. **아동학회지, 34**(1), 71-86.
김애화, 김의정, 김은경, 서선진, 박현, 김은숙 (2011). 특수교육 학회지 동향 분석. **특수교육연구, 18**(2), 49-78.
김애화, 김의정, 김자경, 정대영 (2018). 학습장애, 난독증, 학습부진 (경계선 지능 포함) 및 학습지원대상 학생은 누구이며, 교육적 지원은 이대로 괜찮은가?: 특수교육의 역할과 과제에 대한 소고. **특수교육학연구, 53**(1), 1-21.
김자경, 강혜진, 김기주 (2016). PASS 기반 인지전략 중재가 수학학습장애 위험아동의 수학연산 능력 및 계획기능에 미치는 효과. **특수아동교육연구, 18**(4), 21-37.

김지연 (2017a). 국내 학습장애아 급감 추세와 관련된 요인에 대한 소고. **특수교육학연구, 51**(4), 91-113.

김지연 (2017b). 특수교육대상자 특수학교 배치 현황의 국제 및 국내 분석. **특수아동교육연구, 19**(1), 111-136.

김지원 (2008). 학습장애학생의 수학중재전략 및 관련변인에 관한 메타분석. 대구대학교 석사학위논문.

김화수, 김민정, 박정숙 (2017). 학습장애 학생의 읽기 유창성과 읽기 이해 중재연구에 관한 문헌 분석. **특수교육저널: 이론과 실천, 18**(3), 129-157.

김희진 (2009). 학령 전 아동의 음운처리과정과 초기읽기의 발달 패턴에 관한 연구. 서울대학교 석사학위논문.

남경욱, 신현기 (2008). 인지적 장애를 지난 아동을 위한 보완·대체 의사소통 중재에 대한 메타분석: 단일대상연구를 중심으로. **특수교육학연구, 42**(4), 193-212.

민수진 (2019). 학습장애, 그들은 어디로 갔는가?: 진단, 명칭, 교육에 대한 초등 특수교사 현장 목소리. **학습장애연구, 16**(1), 33-67.

배빛나, 김소희 (2016). 경도장애학생의 기초학습기능향상을 위한 멀티미디어활용 중재효과 메타연구. **학습장애연구, 13**(1), 85-106.

변관석, 신진숙 (2017). 경계선 지능에 관한 국내연구 동향 분석. **특수아동교육연구, 19**(1), 79-109.

서유진, 나경은 (2012). 지적장애 학생을 위한 교과교육 중재연구 분석: 질적지표와 증거기반 중재 기준을 중심으로. **특수아동교육, 14**(1), 435-466.

손승현, 이예다나 (2017). 특수교육요구(Special Needs Education)의 국제적 맥락 고찰. **학습장애연구, 14**(3), 25-47.

신미경, 정평강 (2018). 단일대상 메타분석 연구를 위한 그래프 추출 컴퓨터 프로그램의 신뢰도, 타당도, 사용성 평가. **특수교육, 17**(3), 185-207.

신미경, 채수정, 정평강 (2018). 학습장애 학생들의 문장제 문제해결력 향

상을 위한 전략교수 효과: 단일대상연구 메타분석. **학습장애연구, 15**(3), 203-230.

신재현 (2019a). 교육 사각지대 학습자를 위한 지원 체계. **교원교육소식, 93,** 6-12.

신재현 (2019b). 쓰기학습장애 위험군 선별을 위한 교육과정중심측정(CBM) 쓰기검사의 타당도: 메타분석을 중심으로. **학습장애연구, 16,** 101-122.

신재현 (2019c). 학습장애 및 학습부진 아동을 위한 집중교수(Intensive Intervention)의 재 개념화: 교사에 의한 증거기반교수의 선정과 교수 수정을 중심으로. **학습장애연구, 16,** 93-113.

안예지, 이미지 (2020). 질적지표를 활용한 국내 지적장애 대상 비디오 모델링 중재특성 분석. **지적장애연구, 22**(4), 195-220.

양명희, 김황용 (2002). 개별 피험자 연구에서 사회적 타당도 검사의 중요성. **정서·행동장애연구, 18**(3), 91-111.

여승수. (2010). CBM 읽기 검사의 신뢰도 일반화 연구. 한국특수교육학회 학술대회, 39-52.

여승수, 홍성두, 손지영 (2014). 한국 교육과정중심측정(CBM)의 연구 동향: 측정학적 적합성을 중심으로. **학습장애연구, 11**(3), 187-213.

여승수, 홍성두 (2011). 중재반응모형을 활용한 읽기장애 위험군의 하위유형 분류-잠재성장계층분석을 중심으로. **특수교육 저널: 이론과 실천, 12**(1), 281-298.

이대식 (2015). 학습부진학생의 기초학력 향상을 위한 정부 지원 사업의 특징과 발전 방향. **학습장애연구, 12,** 101-132.

이대식 (2019). 학습 어려움, 어떻게 이해하고 대처할 것인가?: 국내 학습장애 교육 발전을 위한 과제와 방향. **학습장애연구, 16,** 1-32.

이대식, 임건순 (2019). 학습부진학생 지원 체제에 대한 경기 지역 초등교사들의 의견. **교육문화연구, 25**(1), 449-470.

이소현, 박은혜, 김영태 (2000). **교육 및 임상현장 적용을 위한 단일대상연구.** 서울: 학지사.

이승미, 김중훈, 정평강, 송푸름 (2018). 웹기반 CR-CBA 한글해득수준진단검사의 측정학적 적합성. **교육과정평가연구, 21**(4), 125-149.

이애진, 이예다나 (2019). 난독증 학생지원을 위한 지방자치단체 조례 분석. **특수교육학연구, 54**(1), 91-119.

이예다나, 김소영, 길한아, 임보라, 손승현 (2015). 국내 지적장애 학생 대상 읽기 중재 연구의 특징 및 질적 지표 분석. **지적장애연구, 17**(3), 101-129.

전윤희 (2016). 수학학습장애 연구 동향 메타분석. 건국대학교 박사학위논문.

정광조, 홍성두 (2017). 국내 학습부진 및 학습장애 쓰기 중재 연구에 대한 효과 분석: R을 이용한 메타분석을 중심으로. **학습장애연구, 14**(1), 99-122.

정광조, 이대식 (2014). 국내 직접교수 읽기 연구의 증거기반 정도-체계적 연구결과분석 (Systematic Review) 과 메타분석 (Meta-Analysis) 을 중심으로. **아시아교육연구, 15**(4), 265-284.

정동훈 (2009). 국내 보조공학 연구동향 분석: 1998~2008년 특수교육 및 재활 관련 학술지 중심으로. **지체·중복·건강장애연구, 52**(1), 97-124.

조은혜 (2017). 수학 학습부진 중재 프로그램의 연구동향 및 메타분석. 서울교육대학교 석사학위논문.

정평강 (2018). 데이터기반개별화(DBI)의 특성과 학생의 학업성취도에 미치는 효과: 문헌 연구. **학습장애연구, 15**(1), 165-192.

정평강, 신재현, McMaster, K. L. (2016). 데이터기반 교수(Data-Based Instruction)의 쓰기영역 적용 사례: 초등학교 저학년 학습장애 위험 아동을 대상으로. **특수교육, 15**(4), 61-80.

최수미, 유인화, 김동일, 박애실 (2018). 일반교사가 지각하는 교육사각지대 학습자 특성의 구성개념 탐색: CQR-M 을 중심으로. **교육심리연구, 32**(3), 421-442.

최세민, 권택환 (2009). 초등학교 사회과 수업 시 교수적 수정 실태 및

문제점 분석. **지적장애연구, 11**(1), 183-198.
최종근 (2010). 증거기반 특수교육 실제에서 평가 조정의 의의와 원칙. **학습장애연구, 7**(2), 211-239.
최희승 (2020). 질적 지표와 증거기반 중재 기준에 근거한 청각장애 아동의 언어능력 관련 문헌 분석. **한국청각· 언어장애교육연구, 11,** 19-48.
탁효정 (2017). 수학학습부진/학습장애 대상에게 적용한 다양한 중재프로그램의 충실도 및 효과크기 분석. 공주교육대학교 석사학위논문.
한국보건복지인력개발원 아동자립지원단 (2017). **경계선지적기능아동 자립지원체계 연구: 경계선지능아동 자립지원서비스 효과성 보고서**. 보건복지부.
한성희, 남윤석 (2005). 특수교육에서의 단일대상연구 적용의 질적 개선을 위한 문헌 분석. **특수교육학연구, 40**(2), 103-130.
허상 (2008). 초기 수학능력 평가를 위한 수 감각 검사의 타당화 연구. 서울대학교 석사학위논문.
허승준 (2016). 장애 관련 용어와 개념의 재고. **특수아동교육연구, 18**(2), 99-120.
홍성두, 정은주, 김동일 (2006). 수학학습장애 진단을 위한 교육과정중심 측정(Curriculum-Based Measurement: CBM) 검사의 필요성과 전망. **학습장애연구, 3**(1), 1-27.
홍성두, 윤상용, 지은, 김민, 이미지, 김용진, 김주선 (2019). **장애인 고용서비스 판정도구 개발 연구**. 성남: 한국장애인고용공단 고용개발원.
홍이레 (2016). ASD아동을 위한 양육자의 의사소통 중재 연구의 질 및 증거기반의 실제에 대한 평가. 한국특수교육학회 학술대회, 468-477.
Abbott, M., Wills, H., Miller, A., & Kaufman, J. (2012). The relationship of error rate and comprehension in second and third grade oral reading fluency. *Reading Psychology, 33*(1-2),

104-132.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.).* Washington, DC: Author.

Ardoin, S., Christ, T., Morena, L., Cormier, D., & Klingbeil, D. (2013). A systematic review and summarization of the recommendations and research surrounding curriculum-based measurement of oral reading fluency (CBM-R) decision rules. *Journal of School Psychology, 51,* 1 - 18.

Baron, J. (2004). Identifying and implementing education practices supported by rigorous evidence: A user friendly guide. *Journal for Vocational Special Needs Education, 26,* 40-54.

Bast, J., & Reitsma, P. (1998). Analyzing the development of individual differences in terms of Matthew effects in reading: Results from a Dutch longitudinal study. *Developmental Psychology, 34*(6), 1373.

Bender, W. N., & Shores, C. (2007). *Response to intervention: A practical guide for every teacher.* Corwin Press.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis.* John Wiley & Sons.

Bresina, B. C., & McMaster, K. L. (2020). Exploring the relation between teacher factors and student growth in early writing. *Journal of Learning Disabilities, 53*(4), 311-324.

Brossart, D. F., Parker, R. I., Olson, E. A., & Mahadevan, L. (2006). The relationship between visual analysis and five statistical analyses in a simple AB single-case research design. *Behavior Modification, 30,* 531 - 563.

Capizzi, A. M., & Fuchs, L. S. (2005). Effects of curriculum-based measurement with and without diagnostic feedback on teacher

planning. *Remedial and Special Education, 26*(3), 159-174.

Casey, A., Deno, S., Marston, D., & Skiba, R. (1988). Experimental teaching: Changing beliefs about effective instructional practices. *Teacher Education and Special Education, 11,* 123 - 131.

Chard, D. J., Stoolmiller, M., Harn, B. A., Wanzek,J., Vaughn, S., Linan-Thompson, S., & Kame'enui, E. J. (2008). Predicting reading success in a multilevel schoolwide reading model: A retrospective analysis. *Journal of Learning Disabilities, 41,* 174-188.

Chard, D. J., Ketterlin-Geller, L. R., Baker, S. K., Doabler, C., & Apichatabutra, C. (2009). Repeated reading interventions for students with learning disabilities: Status of the evidence. *Exceptional Children, 75*(3), 263-281.

Chauhan, S. (2011). Slow learners: Their psychology and educational programmes. *International Journal of Multidisciplinary Research., 1*(8), 279-289.

Clarke, B., & Shinn, M. R. (2004). A preliminary investigation into the identification and development of early mathematics curriculum-based measurement. *School Psychology Review, 33*(2), 234-248.

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education (5th Ed.).* London and New York: Routledge Falmer.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences.* Academic press.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52,* 219 - 232.

Deno, S. L. (2003). Curriculum-based measures: Development and perspectives. *Assessment for Effective Intervention, 28*(3-4),

3-12.

Deno, S. L., & Mirkin, P. K. (1977). *Data-based program modification: A manual.* Reston, VA: Council for Exceptional Children.

Deno, S. L., Fuchs, L. S., Marston, D., & Shin, J. (2001). Using curriculum-based measurement to establish growth standards for students with learning disabilities. *School Psychology Review, 30*(4), 507-524.

Dennis, M. S., & Gratton-Fisher, E. (2020). Use data based individualization to improve high school Students' mathematics computation and mathematics concept, and application performance. *Learning Disabilities Research & Practice, 35*(3), 126-138.

Dİnçer, S. (2018). Content analysis in for educational science research: Meta-analysis, meta-synthesis, and descriptive content analysis. *Bartin Üniversitesi Egitim Fakültesi Dergisi, 7*(1), 176-190.

Elkins, J. (2000). All empires fall, you just have to know where to push: Antecedent issues for a study of learning difficulties in Australia. *Australian Journal of Learning Disabilities, 5*(2), 4 - 7.

Lembke, E. S., McMaster, K. L., Smith, R. A., Allen, A., Brandes, D., & Wagner, K. (2018). Professional development for data-based instruction in early writing: Tools, learning, and collaborative support. *Teacher Education and Special Education, 41*(2), 106-120.

Espin, C., McMaster, K. L., Wayman, M. M., & Röse, S. (Eds.). (2012). *A measure of success: The influence of curriculum-based measurement on education.* U of Minnesota

Press.

Filderman, M. J., Toste, J. R., Didion, L. A., Peng, P., & Clemens, N. H. (2018). Data-based decision making in reading interventions: a synthesis and meta-analysis of the effects for struggling readers. *The Journal of Special Education, 52*(3), 174-187.

Francis, H. (1992). Patterns of reading development in the first school. *British Journal of Educational Psychology, 62*(2), 225-232.

Fuchs, L. S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review, 33*(2), 188-192.

Fuchs, D., & Fuchs, L. S. (2015). Rethinking service delivery for students with significant learning problems: Developing and implementing intensive instruction. *Remedial and Special Education, 36*(2), 105-111.

Fuchs, L. S., Deno, S. L., & Mirkin, P. K. (1984). The effects of frequent curriculum-based measurement and evaluation on pedagogy, student achievement, and student awareness of learning. *American Educational Research Journal, 21*(2), 449-460.

Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (1989). Effects of instrumental use of curriculum-based measurement to enhance instructional programs. *Remedial and Special Education, 10*(2), 43-52.

Fuchs, L. S., Hamlett, D. F. C. L., & Stecker, P. M. (1991). Effects of curriculum-based measurement and consultation on teacher planning and student achievement in mathematics operations. *American Educational Research Journal, 28*(3), 617-641.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5*(3), 239-256.

Fuchs, D., Mock, D., Morgan, P. L., & Young, C. L. (2003). Responsivenessto-intervention: Definitions, evidence, and implications for the learning disabilities construct. *Learning Disabilities Research & Practice, 18,* 157 - 171.

Fuchs, D., Fuchs, L. S., & Stecker, P. M. (2010). The "blurring" of special education in a new continuum of general education placements and services. *Exceptional Children, 76,* 301 - 323.

Fuchs, D., Fuchs, L. S., & Compton, D. L. (2012). Smart RTI: A nextgeneration approach to multilevel prevention. *Exceptional Children, 78,* 263 - 279.

Fuchs, L. S., Fuchs, D., & Malone, A. S. (2017). The taxonomy of intervention intensity. *Teaching Exceptional Children, 50*(1), 35-43.

Gage, N. A., & Lewis, T. J. (2014). Hierarchical linear modeling meta-analysis of single-subject design research. *The Journal of Special Education*, *48*(1), 3-16.

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional children, 71*(2), 149-164.

Graham, L., & Bailey, J. (2007). Learning disabilities and difficulties: An Australian conspectus—Introduction to the special series. *Journal of Learning Disabilities, 40*(5), 386-391.

Hassiotis, A., Strydom, A., Hall, I., Ali, A., Lawrence-Smith, G., Meltzer, H., Head, J., & Bebbington, P. (2008). Psychiatric

morbidity and social functioning among adults with borderline intelligence living in private households. *Journal of Intellectual Disability Research, 52*(2), 95-106.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*(2), 107-128.

Hedges, L., & Olkin, I. (1985). *Statistical methods for meta-analysis.* London: Academic Press.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta regression with dependent effect size estimates. *Research Synthesis Methods, 1*(1), 39-65.

Heyvaert, M., Maes, B., Van Den Noortgate, W., Kuppens, S., & Onghena, P. (2012). A multilevel meta-analysis of single-case and small-n research on interventions for reducing challenging behavior in persons with intellectual disabilities. *Research in Developmental Disabilities, 33*(2), 766-780.

Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2019). *Cochrane handbook for systematic reviews of interventions.* John Wiley & Sons.

Hintze, J. M., Christ, T. J., & Methe, S. A. (2006). Curriculum-based assessment. *Psychology in the Schools, 43*(1), 45-56.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*(2), 165-179.

Horner, R. H., Swaminathan, H., Sugai, G., & Smolkowski, K. (2012). Considerations for the systematic analysis and use of single-case research. *Education and Treatment of Children, 35*(2), 269-290.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings.* Sage.

Individuals with Disabilities Education Improvement Act of 2004, 20 U.S.C. § 300 *et seq.* (2004).

Jenkins, J. R., Deno, S. L., & Mirkin, P. K. (1979). Measuring pupil progress toward the least restrictive alternative. *Learning Disability Quarterly, 2*(4), 81-91.

Jenson, W. R., Clark, E., Kircher, J. C., & Kristjansson, S. D. (2007). Statistical reform: Evidence based practice, meta analyses, and single subject designs. *Psychology in the Schools*, *44*(5), 483-493.

Jitendra, A. K., Burgess, C., & Gajria, M. (2011). Cognitive strategy instruction for improving expository text comprehension of students with learning disabilities: The quality of evidence. *Exceptional Children, 77(2), 135-159.*

Jung, P. G., McMaster, K. L., Kunkel, A. K., Shin, J., & Stecker, P. M. (2018). Effects of data-based individualization for students with intensive learning needs: A meta analysis. *Learning Disabilities Research & Practice, 33*(3), 144-155.

Katzir, T., Kim, Y., Wolf, M., O'Brien, B., Kennedy, B., Lovett, M., & Morris, R. (2006). Reading fluency: The whole is more than the parts. *Annals of Dyslexia, 56*(1), 51-82.

Kendeou, P., Papadopoulos, T. C., & Spanoudis, G. (2012). Processing demands of reading comprehension tests in young readers. *Learning and Instruction, 22*, 354-367.

Kennedy, C. H. (2005). *Single-case designs for educational research.* Boston: Allyn and Bacon.

Kim, D., Kim, W., Koh, H., Lee, J., Shin, J., & Kim, H. (2014). Exploring students at risk for reading comprehension

difficulties in South Korea: the RTI approach applying latent class growth analysis. *Asia Pacific Education Review, 15*(4), 633-643.

Kim, M. K., Brayant, D. P., & Bryant, B. R., & Park, Y. J. (2017). A synthesis of interventions for improving oral reading fluency of elementary students with learning disabilities. *Preventing School Failure: Alternative Education for Children and Youth, 61*(2), 116-125.

Kirk, S., Gallagher, J., Coleman, M. R., & Anastasiow, N. J. (2011). *Educating Exceptional Children.* Nelson Education.

Kratochwill, T. R., & Levin, J. R. (2014). *Single-case intervention research: Methodological and statistical advances.* American Psychological Association.

Lembke, E. S., McMaster, K. L., Smith, R. A., Allen, A., Brandes, D., & Wagner, K. (2018). Professional development for data-based instruction in early writing: Tools, learning, and collaborative support. *Teacher Education and Special Education, 41*(2), 106-120.

Lemons, C. J., Zigmond, N., Kloo, A. M., Hill, D. R., Mrachko, A. A., Paterra, M. F., Bost, T. J., & Davis, S. M. (2013). Performance of students with significant cognitive disabilities on early-grade curriculum-based measures of word and passage reading fluency. *Exceptional Children, 79*(4), 408-426.

Lyon, G. R., Fletcher, J. M., Shaywitz, S. E., Shaywitz, B. A., Torgesen, J. K., Wood, F. B., Shulte, A., & Olson, R. (2001). Rethinking learning disabilities. *Rethinking Special Education for a New Century,* 259-287.

Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2003). A definition of dyslexia. *Annals of Dyslexia, 53*(1), 1-14.

Marston, D., Muyskens, P., Lau, M., & Canter, A. (2003). Problem solving model for decision making with high incidence disabilities: The Minneapolis experience. *Learning Disabilities Research & Practice, 18*(3), 187-200.

Mattatall, C. (2011). Using CBM to help Canadian elementary teachers write effective IEP goals. *Exceptionality Education International, 21*(1). 61-71.

McDougall, J., Wright, V., & Rosenbaum, P. (2010). The ICF model of functioning and disability: incorporating quality of life and human development. *Developmental Neurorehabilitation, 13*(3), 204-211.

McMaster, K. L., Lembke, E. S., Shin, J., Poch, A. L., Smith, R. A., Jung, P. G., Allen, A. A., & Wagner, K. (2020). Supporting teachers' use of data-based instruction to improve students' early writing skills. *Journal of Educational Psychology, 112*(1), 1-60.

McMaster, K. L., Kunkel, A., Shin, J., Jung, P. G., & Lembke, E. (2018). Early writing intervention: A best evidence synthesis. *Journal of Learning Disabilities, 51*(4), 363-380.

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods, 11*(2), 364-386.

National Center on Intensive Intervention (NCII) at American Institutes for Research. (2013). *Data-based individualization: A framework for intensive intervention.* ERIC Clearinghouse.

Nese, J. F., Park, B. J., Alonzo, J., & Tindal, G. (2011). Applied curriculum-based measurement as a predictor of high-stakes assessment. *The Elementary School Journal, 111*, 608-624.

Odom, S. L., Brantlinger, E., Gersten, R., Horner, R. H., Thompson,

B., & Harris, K. R. (2005). Research in special education: Scientific methods and evidence-based practices. *Exceptional Children, 71*(2), 137-148.

OECD (2003). *Transforming disability into ability: Policies to promote work and income security for disabled people.* Paris: OECD Publishing.

OECD/European Communities (2009). *Students with disabilities, Learning difficulties and disadvantages in the Baltic States, South Eastern Europe and Malta: Educational policies and indicators.* Paris: OECD Publishing.

Oslund, E. L., Elleman, A. M., & Wallace, K. (2020). Factors related to data-based decision-making: Examining experience, professional development, and the mediating effect of confidence on teacher graph literacy. *Journal of Learning Disabilities,* 1 - 13.

Parker, R. I., Vannest, K. J., & Brown, L. (2009). The improvement rate difference for single-case research. *Exceptional Children, 75,* 135 - 150.

Poch, A. L., McMaster, K. L., & Lembke, E. S. (2020). Usability and Feasibility of Data-Based Instruction for Students with Intensive Writing Needs. *The Elementary School Journal, 121*(2), 197-223.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Sage.

Raudenbush, S. W., Bryk, S., Cheong, Y. F., & Congdon, R. (2004). *HLM6: Hierarchical linear and nonlinear modeling.* Chicago: Scientific Software International.

Richards, C., Pavri, S., Golez, F., Canges, R., & Murphy, J. (2007). Response to intervention: Building the capacity of teachers to

serve students with learning difficulties. *Issues in Teacher Education, 16*(2), 55-64.

Robson, C. (2005). *Students with disabilities, learning difficulties and disadvantages: statistics and indicators.* OECD: Organisation for Economic Cooperation and Development,

Romig, J. E., Therrien, W. J., & Lloyd, J. W. (2017). Meta-analysis of criterion validity for curriculum-based measurement in written language. *The Journal of Special Education, 51*(2), 72-82.

Sabornie, E. J., Cullinan, D., Osborne, S. S., & Brock, L. B. (2005). Intellectual, academic, and behavioral functioning of students with high-incidence disabilities: A cross-categorical meta-analysis. *Exceptional Children, 72*(1), 47-63.

Shadish, W. R., Rindskopf, D. M., & Hedges, L. V. (2008). The state of the science in the meta-analysis of single-case experimental designs. *Evidence-Based Communication Assessment and Intervention, 2*(3), 188-196.

Shin, J., Espin, C. A., Deno, S. L., & McConnell, S. (2004). Use of hierarchical linear modeling and curriculum-based measurement for assessing academic growth and instructional factors for students with learning difficulties. *Asia Pacific Education Review, 5*(2), 136-148.

Snyder, S. M., & Ayres, K. (2020). Investigating the usage of reading curriculum-based measurement (CBM-R) to formatively assess the basic reading skills of students with intellectual disability. *Education and Training in Autism and Developmental Disabilities, 55*(1), 60-74.

Stecker, P., Fuchs, L., & Fuchs, D. (2005). Using curriculum based measurement to improve student achievement: Review of research. *Psychology in the Schools, 42,* 795 - 819.

Swanson, H., & Sachse-Lee, C. (2000). A meta-analysis of single-subject-design intervention research for students with LD. *Journal of Learning Disabilities, 33*(2), 114-136.

U. S. Department of Education (2019). *41th Annual report to congress on the implementation of the individuals with disabilities education act..* Washington, DC: Author.

van Kraayenoord, C. E., & Elkins, J. (2004). Learning difficulties in numeracy in Australia. *Journal of Learning Disabilities, 37*(1), 32-41.

Van den Noortgate, W., & Onghena, P. (2003a). Combining single-case experimental data using hierarchical linear models. *School Psychology Quarterly, 18,* 325 - 346.

Van den Noortgate, W., & Onghena, P. (2003b). Hierarchical linear models for the quantitative integration of effect sizes in single-case research. *Behavior Research Methods, Instruments, and Computers, 35,* 1 - 10.

Van den Noortgate, W., & Onghena, P. (2007). The aggregation of single-case results using hierarchical linear models. *The Behavior Analyst Today, 8*(2), 196-208.

Vaughn, S., & Fuchs, L. S. (2003). Redefining learning disabilities as inadequate response to instruction: The promise and potential problems. *Learning Disabilities Research & Practice, 18*(3), 137-146.

Wallace, T., Ticha, R., & Gustafson, K. (2008). *Study of general outcome measurement (GOMs) in reading for students with significant cognitive disabilities: Year 1 (Vol. 27).* RIPM Technical Report.

Wang, S. Y., Cui, Y., & Parrila, R. (2011). Examining the effectiveness of peer-mediated and video-modeling social skills

interventions for children with autism spectrum disorders: A meta-analysis in single-case research using HLM. *Research in Autism Spectrum Disorders* , *5*(1), 562-569.

Wang, S. Y., Parrila, R., & Cui, Y. (2013). Meta-analysis of social skills interventions of single-case research for individuals with autism spectrum disorders: Results from three-level HLM. *Journal of Autism and Developmental Disorders, 43*(7), 1701-1716.

Wayman, M. M., Wallace, T., Wiley, H. I., Ticha, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *The Journal of Special Education, 41, 85-120.*

WHO (2001). *ICF: International classification of functioning, disability and health.* Geneva: Author.

*분석에 사용된 논문

[(준)집단실험설계 연구]

강옥려, 박보영 (2010). 활동중심 수감각 학습이 수학학습장애 위험학생의 수학 성취도에 미치는 영향. **한국초등교육, 20**(2), 215-232.

김동일, 이태수 (2005). 직접교수와 진전도 모니터링이 수학학습부진 및 수학학습장애아동의 기초연산능력 및 발달 패턴에 미치는 효과. **특수교육학연구, 40**(3), 171-189.

김혜영, 강옥려 (2010). 중재반응모형에 의한 수학학습장애학생의 선별과정. **특수교육 연구, 17**(2), 275-299.

이새별, 김광수, 강옥려, 홍성두 (2015). 중재반응모델에 기초한 직접교수가 수학 학습장애아동, 학습부진아동의 연산능력에 미치는 효과. **한국초등교육, 26**(4), 435-458.

이태수 (2010). 감상문 표현활동이 쓰기학습장애 학생의 쓰기능력 향상에 미치는 효과. **학습장애연구, 7,** 1-20.

홍성두, 여승수 (2011). 중재반응모형을 활용한 수학 학습장애 위험군의 하위유형 분류 가능성 탐색-잠재성장계층분석을 중심으로. **특수아동교육연구, 13**(2), 393-413.

[단일대상연구]

고영실, 송재홍 (2019). 수학교과 학습부진아의 학습컨설팅 사례연구. **학교심리와 학습컨설팅, 6**(1), 1-25.

권덕용, 강옥려 (2007). 놀이 활동을 이용한 수학학습이 수학 학습부진아의 연산능력과 수학적 태도 변화에 미치는 효과. **학습장애연구, 4**(2), 71-91.

권미영 (2015). 마음챙김기반 읽기중재가 학습장애 학생의 주의집중행동 및 읽기유창성에 미치는 효과. **특수교육재활과학연구, 54**(2), 1-29.

권혜정 (2014). PASS 읽기향상 프로그램(PREP)에 의한 인지과정 훈련이 경도정신지체학생의 읽기속도 및 읽기정확성에 미치는 효과. 공주대학교 석사학위논문.

길유미 (2015). 옛이야기를 활용한 협력적 글쓰기가 국어 학습부진아의 쓰기 능력에 미치는 영향. 대구교육대학교 석사학위논문.

김경희 (2017). 균형적 언어교육 접근법에 기초한 동시 활동이 국어학습부진아의 읽기 능력에 미치는 영향. 대구교육대학교 석사학위논문.

김동일, 고혜정, 이해린 (2014). 수학학습장애 위험아동의 BASA 1년 사례연구: RTI 체제 기반 한국형 CBM 적용을 중심으로. **특수교육저널: 이론과 실천, 15**(1), 193-213.

김동일, 신혜연, 김희주, 조은정 (2018). 읽기부진 아동을 위한 통합 파닉스 접근법(Synthetic Phonics Approach)에 기반한 음운인식 프로그램 효과성 분석. **특수아동교육연구, 20**(2), 25-50.

김동일, 안예지, 조은정, 최서현 (2020). 학습장애 위험학생 대상 개별화 증거기반 읽기유창성 프로그램 사례연구. **교과교육학연구, 24**(6), 589-601.

김미진, 이재원, 이동철 (2014). 스마트 러닝(Smart Learning)을 이용한 독서지도가 지적장애 학생의 읽기 유창성에 미치는 효과. **특수교육연구, 21**(2), 196-220.

김민경, 이기정 (2015). 역할반전 또래교수가 읽기학습부진아의 읽기유창성 및 학업 자아개념에 미치는 영향. **학습장애연구, 12**(2), 185-201.

김보람 (2011). 수학학습장애위험아 지도 및 진단방법으로서의 중재반응접근(RTI) 효과. 경인교육대학교 석사학위논문.

김은삼 (2015). 초등학교 수학 학습부진학생을 위한 덧셈 뺄셈 직접교수(DI) 프로그램의 적용가능성 탐색. 서울대학교 석사학위논문.

박수현, 이기정 (2015). 동화를 활용한 총체적 언어접근이 읽기학습부진아의 읽기유창성 및 이해능력에 미치는 효과. **한국초등교육, 26**(2), 231-245.

설영기 (2005). 읽기 흥미 과제 교수 Program의 적용을 통한 읽기장애아의 읽기능력 변화과정에 대한 연구. 여수대학교 석사학위논문.

송푸름, 김동일 (2020). 즉시세기(subitizing) 기반 수감각 교수가 초등학교 수학학습장애 위험학생의 수감각과 수세기에 미치는 영향. **학습장애연구, 17**(1), 127-157.

신수정, 강옥려 (2018). 명시적 교수에 기반한 통합파닉스 훈련이 난독증 위험학생의 단어재인, 읽기유창성, 철자쓰기에 미치는 효과. **학습장애연구, 15**(3), 103-134.

심말옥, 박현옥 (2015). E-NIE 프로그램을 활용한 국어과 수업이 지적장애학생의 읽기 유창성 및 쓰기 능력에 미치는 영향. **발달장애연구, 19**(3), 123-146.

안미선 (2012). 컴퓨터 보조 학습이 중학교 지적장애 학생의 덧셈·뺄셈 수행능력에 미치는 효과. 단국대학교 석사학위논문.

안성진 (2018). 읽기부진 초등학생의 읽기성취 향상을 위한 비계설정 중재 프로그램의 효과. 서울대학교 석사학위논문.

유은경 (2016). 국어교과 차별화 교수가 국어 학습부진아의 읽기 능력

향상에 미치는 효과. 대구교육대학교 석사학위논문.
유종완 (2010). 독서 감상 쓰기 표현활동이 중학교 학습부진 학생의 쓰기 능력 및 유창성 향상에 미치는 효과. 전남대학교 석사학위논문.
이은옥 (2015). 또래교수를 이용한 감상문 쓰기 교수가 학습장애 위험학생의 쓰기 능력 향상에 미치는 효과. 전남대학교 석사학위논문.
이진영 (2015). 스토리 리텔링 활동이 ADHD 위험군 아동의 읽기능력에 미치는 영향. 대구교육대학교 석사학위논문.
이혜진, 이태수 (2019). 그래픽조직자를 활용한 언어경험접근법이 경도장애학생의 쓰기 능력에 미치는 효과. **특수교육연구, 26**(1), 87-108.
정광조, 이효자 (2009). 중재반응모형의 적용 가능성 탐색 연구. **특수교육학연구, 44**(2), 313-339.
정채미 (2015). 학습장 정리하기 활동이 국어학습부진아의 쓰기 능력에 미치는 영향. 대구교육대학교 석사학위논문.
조경화 (2008). Readers Theater 프로그램이 읽기저성취 아동의 읽기유창성과 읽기이해에 미치는 영향. 서울대학교 석사학위논문.
최은주 (2006). 읽기전략훈련이 읽기장애아의 읽기능력 및 태도에 미치는 효과. 대구대학교 석사학위논문.
최종근 (2008). 교육과정중심측정에서 목표 설정의 의미와 중재 효과. **아시아교육연구, 9**(2), 89-112.
한연숙 (2014). 독서토론 활동이 쓰기 학습부진아의 쓰기능력에 미치는 영향. 대구교육대학교 석사학위논문.
황애희 (2012). 읽기부진 아동의 읽기 능력 향상을 위한 발음중심-직접교수와 유창성 훈련 프로그램의 효과. 덕성여자대학교 석사학위논문.

부록

[부록 1] (준)집단실험설계 질적 지표 평가 루브릭

연구대상자	1(미충족)	2(부분충족)	3(충족)
연구대상자의 장애, 어려움에 대한 정보를 충분히 제공하였는가?	연구대상 관련 정보를 제공하지 않음	연구대상 관련 제한된 정보를 제공함	연구대상 관련 충분한 정보를 제공함
집단 간 연구대상자의 동질성이 확보되었는가?	동질성 확보를 위한 과정을 기술하지 않음	동질성 확보를 위한 과정을 기술하였으나, 적절하지 않거나 충분하지 않음	동질성 확보를 위한 과정이 적절하고, 충분히 기술함
중재제공자(교사)에 대한 정보를 충분히 제공하고, 조건 간 중재제공자의 동등성을 확보하였는가?	중재제공자를 기술하지 않음	중재제공자를 기술하였으나, 중재제공자 및 조건 간 동등성에 대한 정보가 충분하지 않음	중재제공자 및 조건 간 동등성에 대한 정보를 충분히 기술함
중재 실행 및 비교집단 기술	**1(미충족)**	**2(부분충족)**	**3(충족)**
중재가 명백하고 자세하게 기술되었는가?	중재에 대하여 기술하지 않음	중재에 대하여 기술하였으나, 충분하지 않거나 구체적이지 않음	중재에 대하여 충분하고 구체적으로 기술함
중재과정 충실도가 기술되고 측정되었는가?	중재 충실도를 측정 또는 기술하지 않음	중재 충실도를 간접적으로 측정하고 기술함	중재 충실도를 직접적으로 측정하고 구체적인 과정을 기술함
비교집단에 제공된	비교집단에	비교집단에	비교집단에

교수, 서비스 등 특성을 기술하였는가?	제공된 중재를 기술하지 않음	제공된 중재를 기술하였으나, 충분하지 않거나 구체적이지 않음 (단일집단설계의 경우, 2점 부여)	제공된 중재에 대하여 충분하고 구체적으로 기술함
결과 측정도구	**1(미충족)**	**2(부분충족)**	**3(충족)**
중재에 적절한 측정도구를 사용하고, 일반화 수행의 도구 간 균형을 위해 여러 측정도구를 사용하고 있는가?	적절한 측정도구가 사용되지 않음	측정도구가 적절하나, 단일한 측정도구를 사용함	측정도구가 적절하고, 일반화된 수행을 위한 복수의 측정도구를 사용함
결과는 적절한 시기에 측정되었는가?	결과 측정 시기를 기술하지 않음	결과 측정 시기를 기술하였으나, 적절하지 않음	결과 측정 시기가 적절하고 구체적으로 기술함
자료 분석	**1(미충족)**	**2(부분충족)**	**3(충족)**
연구문제에 적합한 분석기법이 사용되고, 분석단위가 적절히 설정되었는가?	분석방법을 기술하지 않음	분석방법을 기술하였으나, 분석방법 및 단위가 연구문제에 적절하지 않음	분석방법 및 단위가 연구문제에 적절하고 구체적으로 기술함
추리통계 뿐 아니라 효과크기 계산을 포함하였는가?	효과크기를 계산하지 않음	효과크기를 계산하였으나 해석이 적절하지 않거나, 결과의 변화 크기를 대안적 방법으로 나타냄	효과크기를 계산하고 적절히 해석함

[부록 2] 단일대상설계 질적 지표 평가 루브릭

연구대상자 및 연구환경	1(미충족)	2(부분충족)	3(충족)
연구대상에 대한 정보를 구체적으로 서술하였는가? (예: 나이, 성별, IQ, 장애유무, 진단명, 사회경제적 수준, 학습수준 평가결과 등)	연구대상 관련 정보를 제공하지 않음	연구대상 관련 제한된 정보를 제공함	연구대상 관련 충분한 정보를 제공함
연구대상자 선별기준 및 선정절차가 반복 가능하도록 정확하게 제시되었는가?	선별기준 및 선정절차를 기술하지 않음	선별기준 및 선정절차를 기술하였으나, 적절하지 않거나 제한된 정보를 제공함	선별기준 및 선정절차가 적절하고 충분히 기술됨
중재제공자에 대한 정보를 구체적으로 서술하거나 (예: 경력, 교사자격증, 교육수준, 나이, 성별, 인종, 중재친숙도 등), 조건 간 중재제공자의 동등성에 대해 기술하였는가?	중재제공자를 기술하지 않음	중재제공자를 기술하였으나, 중재제공자 및 조건 간 동등성에 대한 정보가 충분하지 않음	중재제공자 및 조건 간 동등성에 대한 정보를 충분히 기술함
연구가 수행된 물리적 환경이 구체적으로 기술되었는가?	연구환경을 기술하지 않음	연구환경 관련 제한된 정보를 제공함	연구환경 관련 충분한 정보를 제공함
종속변수	**1(미충족)**	**2(부분충족)**	**3(충족)**
종속변수를 조작적으로 명확하게 기술하고 있는가?	종속변수를 기술하지 않음	종속변수를 기술하였으나, 구체적이지 않거나 조작적으로	종속변수를 구체적이고 조작적으로 기술함

		기술하지 않음	
모든 종속변수에 대하여 양적으로 측정가능한 지표를 산출하도록 측정되었는가?	종속변수에 대해 양적으로 측정가능한 지표를 산출하지 않음	일부 종속변수에 대해 양적으로 측정가능한 지표를 산출함	모든 종속변수에 대해 양적으로 측정가능한 지표를 산출함
측정이 타당하며 반복 가능하도록 측정과정에 대해 명확하게 기술하였는가?	측정과정을 기술하지 않음	측정과정을 기술하였으나, 타당하지 않거나 제한된 정보를 제공함	측정과정이 타당하고 반복 가능함
측정이 충분한 빈도수로 반복적으로 수행되었는가?	반복 측정을 실시하지 않음	반복 측정을 하였으나, 빈도수가 충분하지 않음	반복측정이 충분한 빈도수로 실시됨
각각의 종속변인에 대해 측정 신뢰도 또는 관찰자간 일치도 자료가 수집되었으며, 최소한의 기준에 도달하였는가?	측정 신뢰도를 측정 또는 기술하지 않음	측정 신뢰도를 측정 또는 기술하였으나, 최소 기준을 충족하지 않음	측정 신뢰도를 측정, 기술하고, 최소 기준을 충족함
독립변수	**1(미충족)**	**2(부분충족)**	**3(충족)**
독립변수가 반복 가능하도록 정확하게 기술되었으며 자세한 정보를 제공하였는가? (예: 중재의 개념적 기초, 교수 자료 및 과정, 중재기간, 회기 시간, 교사의 행동 및 언어 등)	독립변수를 기술하지 않음	독립변수를 기술하였으나, 구체적이지 않거나 제한된 정보를 제공함	독립변수를 구체적이고 충분히 기술함
연구자 통제 하 독립변수가 조작되고 실험통제에 대해 정확하게 기술하였는가?	연구자의 통제 노력 없이 독립변수가 제공됨	연구자의 통제 하에 독립변수가 조작되었으나, 통제가 충분하지 않음	조건에 대한 무작위할당이 이뤄졌거나, 연구자의 통제 노력이 충분하고

			구체적으로 기술됨
중재 충실도에 대한 측정과 명확한 기술이 이루어졌는가?	중재 충실도를 측정 또는 기술하지 않음	중재 충실도를 간접적으로 측정하고 기술함	중재 충실도를 직접적으로 측정하고 구체적으로 기술함
기초선	**1(미충족)**	**2(부분충족)**	**3(충족)**
기초선 기간에 종속변수가 자주 측정되었고 앞으로의 수행을 예측할 수 있도록 중재 전 안정화되었는가?	기초선 기간을 설정하지 않거나 종속변수가 측정되지 않음	종속변수가 낮은 빈도수로 측정되거나, 중재 전 수행이 안정화되지 않음	종속변수가 빈번하게 측정되고, 중재 전 수행이 안정화됨
기초선 기간에 대해 반복 가능하도록 구체적인 정보를 제시하였는가? (예: 절차, 환경 등)	기초선 기간을 설정하지 않거나 충분히 기술하지 않음	기초선 기간을 충분히 기술하였으나, 구체적이지 않거나 반복가능하지 않음	기초선 기간을 구체적이고 반복가능하게 기술함
실험통제/내적타당도	**1(미충족)**	**2(부분충족)**	**3(충족)**
실험효과가 3번 이상의 측정치에서 충분히 증명되었는가?	실험효과에 대해 기술하지 않음	실험효과 1~2번의 측정치에서 증명함	실험효과를 3번 이상의 측정치에서 증명함
내적 타당도 위험요소를 대부분 통제하였는가?	내적 타당도 위험요소를 통제하지 않음	내적 타당도 위험요소 중 일부만을 통제함	내적 타당도 위험요소 중 대부분 혹은 전부를 통제함
결과가 증명되었는가? (예: 결과 또는 수준의 변화)	결과 또는 수준의 변화가 나타나지 않음	결과 또는 수준의 변화가 나타났으나, 제한된 정보를 제공함	결과 또는 수준의 변화 패턴이 증명됨 (시각적 그래프 등)

		(경향성, 수준, 변동성 등)	
외적타당도	**1(미충족)**	**2(부분충족)**	**3(충족)**
독립변수의 효과가 대상자 간, 환경 간, 검사도구 간 반복되었는가?	독립변수의 효과 반복을 시도하지 않음	독립변수의 효과 반복을 일부 시도하였으나, 충분하지 않음	독립변수의 효과 반복을 대상자 간, 환경 간, 검사도구 간 시도함
사회적타당도	**1(미충족)**	**2(부분충족)**	**3(충족)**
종속변수가 사회적으로 중요한가?	사회적으로 중요하지 않음	-	사회적으로 중요함
중재를 통한 종속변수의 변화 정도가 사회적으로 중요한가? (예: 평균수준, PND 등)	변화의 정도가 사회적으로 중요하지 않음	변화 정도가 일부 사회적으로 중요하거나, 변화 정도를 구체적으로 기술하지 않음	변화 정도가 사회적으로 중요하고, 변화 정도를 구체적으로 기술함
독립변인 실행의 실용성, 비용효과성에 대한 설명을 제공하였는가? (예: 사회적 타당도 자료 제공, 수용성, 중재적용가능성, 효과성과 계속적인 사용에 대한 설명)	독립변인 실행 관련 정보를 제공하지 않음	독립변인 실행 관련 정보를 제공하였으나 충분하지 않고, 사회적 타당도를 기술하지 않음	독립변인 실행 관련 정보를 충분히 제공하고 사회적 타당도를 기술함
독립변수의 독립적인 수행 특성에 대한 정보를 제공하였으며. 일반적인 상황 속 일반적인 중재자에 의해서도 향상 가능한가?	일반적 상황 속 독립적 수행에 대해 기술하지 않음	일반적 상황 속 독립적인 수행에 대해 기술하였으나 충분하지 않음	일반적 상황 속 독립적인 수행에 대해 충분히 기술함

[부록 3] 단일대상설계 연구별 IRD

저자	성별	연령(학년)	학생유형	종속변수	종속변수 세부	robust IRD
심말옥, 박현옥 (2015)	여	16(고1)	지적장애	쓰기		1
	여	16.7(고1)				0.9
	여	17.2(고2)				0.9
	여	16(고1)	지적장애	읽기유창성		1
	여	16.7(고1)				0.67
	여	17.2(고2)				0.62
권혜정 (2014)	여	17.7(고2)	지적장애	읽기유창성	읽기속도	0.4
	여	19.3(고2)				0.6
	남	20.3(고3)				0.8
조경화 (2008)	남	9(초2)	저성취	읽기유창성	읽기속도	1
	남	9(초3)				1
	남	9(초3)				1
	남	9(초2)	저성취	읽기이해	내용단위 및 구성요소 점수 평균	1
	남	9(초3)				1
	남	9(초3)				0.8
최종근 (2008)	여	8(초2)	저성취	연산유창성		0.96
유은경 (2016)	남	11.3(초6)	학습부진	읽기유창성		1
	남	11.5(초6)				1
	여	11.6(초6)				1
김경희 (2017)	남	7.6(초2)	학습부진	읽기유창성		0.75
	남	7.3(초2)				0.75
	남	7.3(초2)				1
이혜진, 이태수 (2019)	여	11	지적장애	쓰기유창성	완전한 문장 수	0.48
	남	11				0.44
	여	12				0.88
권덕용,	남	9(초4)	학습부진	수연산		0.25

강옥려 (2007)	남	10(초4)				0.5
	남	9(초4)				0.4167
유종완 (2010)	-	중1	학습부진	쓰기유창성		1
	-	중1				1
	-	중1				1
한연숙 (2014)	여	9(초3)	학습부진	쓰기유창성	정량적 점수	1
	여	9.3(초3)				0.94
	남	9.1(초3)				0.69
박수현, 이기정 (2015)	남	초3	학습부진	읽기유창성		0.75
	여	초3				0.94
	여	초3				0.88
	남	초3	학습부진	읽기이해	내용단위 및 구성요소 평균	0.88
	여	초3				0.94
	여	초3				0.94
이은옥 (2015)	남	9.1(초3)	학습장애 위험	쓰기유창성		0.87
	남	8.7(초3)				0.42
	남	9.4(초3)				0.4
권미영 (2015)	남	11.5(초5)	학습장애	읽기유창성	읽기속도	1
	여	11.3(초5)				1
	남	11.8(초5)				1
신수정, 강옥려 (2018)	남	7.8(초2)	난독증 위험	초기읽기	단어재인	1
	남	7.4(초2)				1
	여	7.9(초2)				1
	남	7.8(초2)	난독증 위험	읽기유창성		1
	남	7.4(초2)				1
	여	7.9(초2)				0.85
	남	7.8(초2)	난독증 위험	철자쓰기		1
	남	7.4(초2)				1

	여	7.9(초2)				0.75
김동일, 고혜정, 이해린 (2014)	여	8.6(초3)	학습장애 위험	수연산	RTI 1단계 대비 2단계 성취	0.75
	남	8.2(초3)				0.75
김보람 (2011)	-	초3	학습장애 위험	수연산	RTI 1단계 대비 2단계 성취수준	0.5
	-	초3				0.5
	-	초3				0.5
	-	초3				0.33
	-	초3				0.67
	-	초3				0.33
고영실, 송재홍 (2019)	여	11(초4)	학습부진	수연산		0.8
김미진, 이재원, 이동철 (2014)	여	10.3(초4)	지적장애	읽기유창성	읽기속도	0.92
	남	11.6(초6)				1
이진영 (2015)	남	7.3(초2)	ADHD	읽기유창성		1
	남	7.6(초2)				1
	남	7.6(초2)				0.75
김민경, 이기정 (2015)	남	9.3(초3)	학습부진	읽기유창성		0.81
	남	9.8(초3)				0.88
	여	9.8(초3)				0.75
길유미 (2015)	남	7.5(초2)	학습부진	쓰기유창성	정량적 점수	0.83
	남	7.4(초2)				0.92
	남	7.3(초2)				0.83
설영기 (2005)	남	7.1(초2)	학습장애	읽기유창성		0.80
	여	9.4(초3)				0.80
	남	9.9(초3)				0.80
김동일	남	초1	학습부진	음운인식	하위	0.67

외 (2018)	남	초2			과제별 평균 점수, 사전 대비 진전도 성취수준	1
	여	초2				0.67
	남	초3				1
	여	초3				1
황애희 (2012)	남	8.6(초3)	학습부진	읽기유창성		1
	남	8.6(초3)				0.39
	여	8.7(초3)				0.89
	여	9.1(초3)				0.56
	여	9.2(초3)				1
안성진 (2018)	여	초3	학습부진	읽기유창성		0.85
	남	초4				0.45
	남	초4				0.78
	여	초3				0.88
	여	초3	학습부진	읽기이해	사전 대비 진전도 성취수준	0.67
	남	초4				1
	남	초4				1
	여	초3				1
최은주 (2006)	남	8.6(초3)	학습부진 장애	읽기유창성		1
	여	8.7(초3)				1
	남	8.7(초3)				0.8
	남	7.1(초3)				1
정광조, 이효자 (2009)	남	중1	학습부진	읽기유창성		1
	여	중1				1
	남	중2				0.63
	여	중1	학습장애			1
송푸름, 김동일 (2020)	남	7.9(초1)	학습장애 위험	수감각		1
	여	7.3(초1)				0.86
	여	6.6(초1)				1
김은삼	여	7(초2)	학습부진	연산유창성		1

(2015)	남	8(초2)				0.86
	남	7(초2)				0.86
안미선 (2012)	남	14.3(중2)	지적장애	기초연산		1
	남	15.2(중3)				1
	남	15.5(중3)				1
정채미 (2015)	남	9.6(초4)	학습부진	쓰기유창성	정량적 점수	1
	여	9.5(초4)				1
	남	9.3(초4)				1
김동일 외 (2020)	남	초3	학습장애 위험	읽기유창성	정확음절수	0.67
	여	초4				0.67
	남	초3				1
Robust IRD 평균 = 0.84						

Abstract

A Single-Subject Meta-Analysis of the Effects of Data-Based Instruction for Students with Learning Difficulties in Korea

Seohyeon Choi
Major in Special Education
Department of Education
The Graduate School
Seoul National University

In general school classrooms, there are students who have severe difficulties in acquiring and using basic learning skills for various reasons. These students include those with learning disabilities, dyslexia, emotional and behavioral problems, below-average or

borderline intelligence, multicultural backgrounds, environmental deficits, and insufficient educational opportunities. 'Learning difficulties' is an umbrella term for heterogeneous learners who require additional support due to the significant underachievement and difficulties in the basic academic areas (i.e., reading, writing, and mathematics). It is quite difficult to clearly distinguish these subgroups in the educational field. Among students with learning difficulties, some are identified as having special educational needs and can receive support. However, others, which is a large number, are just left in the classroom without being provided with adequate support to resolve their difficulties.

There has been increasing concern that difficulties of these students should be recognized as 'special educational needs' as well and appropriate interventions should be provided prior to diagnosis, identification, and classification. In Korea, as social interest in diverse students with learning difficulties has gradually increased, some legal-institutional foundations and support systems has been established and improved. Nevertheless, the number of students with learning difficulties in general classrooms is increasing and many of them are reported to be left in blind spots of education.

Data-based instruction (DBI) is an educational approach that is suggested as effective to enhance achievements of struggling learners, particularly for those who did not respond to intensive intervention in usual ways. DBI is defined as a series of successive and systematic procedures in which students' data are used to determine instructional changes. In DBI, teachers monitor and measure the progress of each student frequently while providing research-based interventions. Whether instructional changes are required, when and how it should be done are determined based on the collected data. For progress

monitoring, curriculum-based measurement (CBM) is used. DBI can be applied in a multi-tiered systems of support for academic and behavioral intervention.

If a specific intervention or instructional strategy is to be actively implemented in educational practices, it must be proved that it is an evidence-based practice. To determine whether the specific intervention can be considered as an evidence-based practice, both the quality of individual studies in terms of research design and method and the overall effectiveness of the intervention from different study results should be examined. In Korea, although research on DBI has been steadily conducted especially since the development of the Basic Academic Skills Assessment (BASA), there is a lack of study that tries to identify the level of evidence-based practice of DBI research. Therefore, this study aims to synthesize the DBI research so far in Korea by analyzing the effectiveness of DBI for school-aged students with learning difficulties via meta-analysis and evaluating the quality of the research.

In this study, a total of 32 single-subject design studies were used. Group experimental or quasi-group design studies, which were only six were briefly reviewed. The single-subject design studies were analyzed based on two specific methods. First, a qualitative analysis was conducted using the quality indicators suggested by Horner et al. (2005). Second, meta-analysis was conducted to quantitatively synthesize the effects of DBI on school-aged students with learning difficulties. Data standardization and effect size calculation were based on the method suggested by Van den Noortgate & Onghena (2003a; 2003b; 2007). Then, a two-level multilevel meta-analysis was implemented using the HLM program. In this way, the mean effect size was calculated at the level of study participants. Then, the

heterogeneity between the effect sizes and some predictor variables affecting the heterogeneity were examined.

The main results of this study are as follows. First, it was identified that research on DBI for students with learning difficulties have been conducted constantly in Korea since 2000s. However, a significant number of studies did not report information sufficiently regarding the specific standards and procedures used for instructional change and data-based decision making. Second, the results of the qualitative analysis showed that the overall quality of DBI research was sufficient. Nevertheless, it was found that some aspects such as description of interventionist, comparability of interventionists across conditions, fidelity of intervention, manipulation of independent variable, and reliability of data collection were not reported explicitly, which can lower the reliability of the study results. Third, as the result of multi-level meta-analysis, the mean effect size of DBI was 1.34 and statistically significant. Fourth, there was significant variance across participants in effect sizes. The results from the conditional model showed that exceptionality type, the number of sessions, and the length of each session were significantly accountable for the variability of effect sizes. Fifth, while the provision of support for teachers has been constantly reported to increase the effect of DBI, there were only few studies that included teacher support in Korea. Based on these findings, implications and study limitations were discussed.

keywords : Data-based instruction (DBI), Learning difficulties, Quality indicators, Meta-analysis, Evidence-based practice
Student Number : 2019-29606

Summary

1. Introduction

In general school classrooms, there are groups of students who have severe difficulties in acquiring and using basic learning skills. These are very diverse and heterogeneous groups, and may be affected by a variety of factors, including learning disabilities, dyslexia, emotional or behavioral problems, below-average or borderline intelligence, multicultural backgrounds, environmental deficits, and insufficient educational opportunities. The term 'learning difficulties' refers to the large group of students who exhibit severe problems with learning and need extra assistance with schooling (van Kraayenoord & Elkins, 2004). This generic term reflects the willingness to provide proper intervention and educational services to diverse students with or without internal causes, who exhibit severe difficulties in basic academic areas. It is also related to the OECD model of students with "special educational needs", which encompasses disability, difficulties and social disadvantage (OECD, 2003; Robson, 2005). In this comprehensive and resource-based approach, heterogeneous groups of students can be inclusively involved (Robson, 2005), and they can be served before referral to special education eligibility.

Data-based instruction (DBI) is suggested as an effective approach to enhance academic outcomes of students with severe learning difficulties. DBI, which is also called 'data-based instruction' (Fuchs, Fuchs, & Stecker, 2010), 'data-based individualization' (NCII, 2013), 'data-based program modification' (Deno & Mirkin, 1977), or 'experimental teaching' (Casey et al., 1988), is defined as a series of

successive and systematic procedures in reading, mathematics, or behavior where students' data of achievement are used to determine when and how to intensify and modify interventions (National Center on Intensive Intervention, 2013). To be specific, the procedures of DBI involve the following steps: (a) identifying a student's current level of performance, (b) establishing a long-term academic goal, (c) implementing quality evidence-based intervention with fidelity while monitoring one's progress frequently, (d) using data-based decision rules in order to determine whether instructional changes would be needed, (e) establishing a tentative hypothesis about the student's specific needs and implementing changes in instruction based on the previous hypotheses, (f) evaluating the effectiveness of the instructional changes based on progress monitoring data, and (g) repeating these procedures until the student achieves the academic goal (Jung et al., 2018; Lembke et al., 2018). It can be implemented either as part of special education services for students with disabilities or as part of multi-tiered systems in general education such as an approach of Responsiveness-to-Intervention (RTI) (Fuchs, Fuchs, & Compton, 2012; NCII, 2013). A strong evidence base supports the efficacy of DBI for students with learning difficulties (Stecker, Fuchs, & Fuchs, 2005; Fuchs, Fuchs, & Hamlett, 1989; Poch, McMaster, & Lembke, 2020). Moreover, DBI has positive influence on teachers' instructional planning, leading them to make more specific plans, to make instructional adaptations more frequently, and to identify the targeted skills more appropriately (Capizzi & Fuchs, 2005; Fuchs, Deno, & Mirkin, 1984).

It is important to use a reliable measurement that is sensitive to the struggling students' growth over a short period of time in order to make ongoing decision in DBI (Jung et al., 2018). In this purpose,

curriculum-based measurement (CBM) is used to decide whether to raise the goal of instruction, keep instruction as-is, or change instruction (McMaster et al., 2020). CBM, which was first developed by Stanley Deno and his colleagues at the University of Minnesota, is a standardized measurement for assessing students' academic competence and progress in the basic academic domains (Deno, 1985). It is inexpensive and efficient to implement and has high adequacy in terms of validity and reliability (Romig, Therrien, & Lloyd, 2017). Moreover, as a 'general outcome measure', the scores of CBM are considered to represent an individual's generalized level of the corresponding domain (Deno et al., 2001).

In Korea, the concept of DBI was introduced and has been applied for teaching school-aged students with learning difficulties, especially since 2000 when the Korean CBM was developed as the name of Basic Academic Skills Assessment (BASA) (Kim, 2000; Espin et al., 2012). According to Yeo, Hong, & Son (2014) who reviewed the trend of CBM research in Korea from 1999 to 2014, BASA was found to be the most widely used CBM tool in Korea. So far, BASA has been developed in eight basic academic areas: reading, writing, mathematics, math word problems, vocabulary, reading comprehension, early mathematics, and early literacy. Since BASA includes a number of equivalent tests, it can be conducted at regular basis for progress monitoring in the process of DBI. For data-based decision rules, it recommends using the data-point method (i.e., to make instructional change when more than 4 successive scores below or more than 3 successive scores above the target line) and the slope-based method (i.e., to compare the slope of the progress line with the slope of the target line).

In educational practice, general and special teachers are required to

use interventions which were scientifically proven. In the field of special education, there have been constant efforts to identify and implement evidence-based practices (Baron, 2004). In order to examine whether the specific intervention and the studies that implemented it have some scientific evidence, it is first necessary to evaluate the quality of the individual studies in terms of study design and method (Odom et al., 2005). For evaluating the quality of studies in special education field, the quality indicators (QIs) for four types of research methodologies (i.e., group experimental, correlational, single subject, and qualitative designs) were developed by the task force team at the Council for Exceptional Children's (CEC) (Odom et al., 2005). Specifically, Gersten et al. (2005) identified QIs for group experimental and quasi-experimental studies, and Horner et al. (2005) identified QIs for single-subject designs. Second, it should be proved in a number of studies that the effectiveness of the intervention is considerably large. In this purpose, meta-analysis is widely used to statistically synthesize the results from more than two separate studies (Higgins et al., 2019).

In the present study, it aims to analyze the overall effectiveness of DBI for school-aged students with learning difficulties and the quality of research in Korea. The present study attempts to collect and synthesize only single-subject design studies. Specifically, it examines the quality of the studies using the QIs suggested by Horner et al. (2005). In addition, it synthesizes the effectiveness of those interventions via meta-analysis and identifies the potential variables to explain the variances of the effect sizes. As of now in Korea, there is a lack of research that specifically synthesizes the effectiveness of DBI. Although Yeo, Hong, & Son (2014) synthesized CBM research in Korea, their research questions were focused on

CBM and the aspects of technical adequacy. Moreover, it was not a meta-analytic review but a narrative one. Another literature review conducted by Jung (2018) was focused on the theme of DBI, but only international studies were included. Therefore, the present study poses the following research questions:

RQ1: What are the characteristics and quality of studies on DBI implemented for students with learning difficulties?

RQ2: What is the average effect of DBI on basic academic skills for students with learning difficulties?

RQ3: To what extent have participant-, intervention-, DBI-related variables influence on the effectiveness of DBI for students with learning difficulties?

2. Methods

2.1. Search Procedure

A comprehensive literature search was conducted using the Korean electronic databases RISS (Research Information Sharing Service) and KISS (Korean studies Information Service System), and Nurimedia DBpia in order to retrieve studies written in Korean. Since the number of research on DBI for students with learning difficulties was not large, no restriction was set regarding dates of publication. For this search, keywords for students with learning difficulties ("learning difficulties", "learning disabilities", "at-risk", "low achievement", and "underachievement"), keywords for basic academic skills ("basic academic", "basic learning", "reading", "writing", and "math"), keywords related to DBI ("data-based instruction", "evidence-based instruction", "data-based individualization", and "curriculum-based

measurement"), and keywords for BASA ("Basic Academic Skills Assessment", and "BASA") were used in various combinations, and Korean terms were used for the search. Furthermore, reference lists both from prior syntheses and the included studies were reviewed to encompass any possible studies to be analyzed. By including the keywords for BASA, it reduced the likelihood to miss the studies that did not use the term "DBI" explicitly but were grounded on DBI using BASA. This database search yielded 422 studies after duplicates removed. From these, a total of 390 studies were excluded from the final analyses based on the following inclusion and exclusion criteria.

2.2. Inclusion and Exclusion Criteria

Among the preselected studies, only studies which met all the following criteria and had full text accessible were included in the analysis. First, studies with students with learning difficulties as participants were included. Terms such as "students with/at-risk of learning disabilities", "struggling learners", "low achievers", and "under-achievers" were considered to fall within the category of students with learning difficulties. Studies for participants with intellectual disabilities were included in this study. Second, participants were school-aged children from six to eighteen. Third, studies that implemented intervention on basic academic skills (reading, writing, and math) and had dependent variables for those skills were included. There was no restriction in the type of intervention as long as basic academic skills were set as dependent variables. In several studies of which aims were beyond to determine the effectiveness of intervention (e.g., latent class analysis of struggling learners, examination of the applicability of RTI approach,

or screening of students with learning disabilities), they were accepted when they had implemented the procedures of DBI on basic academic skills. Fourth, studies that conducted intervention following the principal rules of DBI (i.e., frequent progress monitoring and the use of data for ongoing instructional decision making) were included. Fifth, only the studies that used BASA as CBM tools were included for the analysis, which is the standardized and most frequently used monitoring tool and has high technical adequacy. Other researcher- and teacher-generated CBM were excluded, as they had been pointed out to have a lack of evidence of validity and reliability (Jung et al., 2018; Yeo, Hong, & Son, 2014). Sixth, studies of single-subject designs providing enough quantitative statistics to calculate effect sizes were accepted. Figure 1 summarizes these procedures of literature selection according to the PRISMA 4-phase flow diagram.

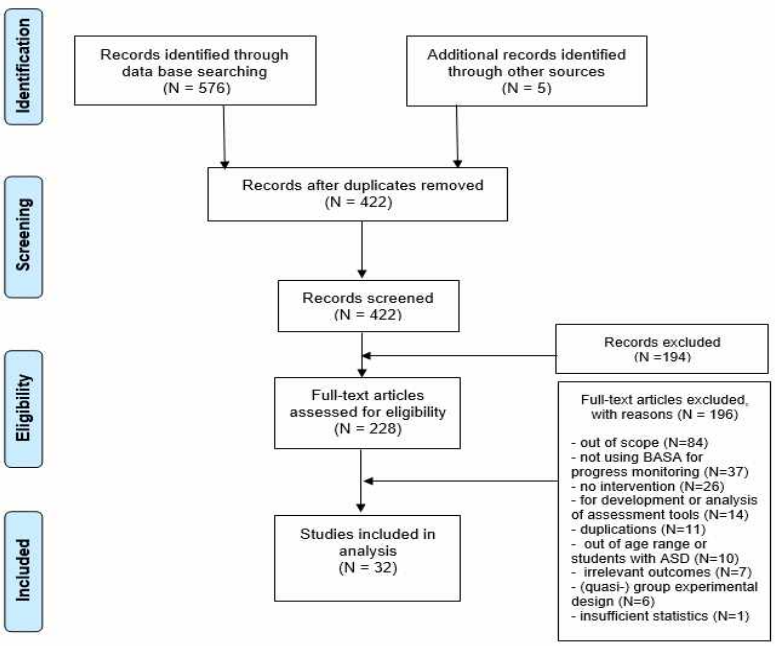


Figure 1. PRISMA flowchart for data collection

2.3. Coding Procedures

2.3.1. Coding System for Quality Indicators

QIs suggested by Horner et al. (2005) were used to examine each study. However, it has been pointed out in the previous studies that these QIs have strict criteria and that evaluating in a dichotomous way is quite difficult (Jitendra, Burgess, & Gajria, 2011; Ahn & Lee, 2020; Lee et al., 2015). Thus, while the present study is primarily based on the QIs of Horner et al. (2005), it evaluated the quality level of each indicator in a continuum using the 3-point Likert scale. For QI rating, it referred to the rubric suggested by Chard et al. (2009) which was based on 4-point Likert scale. The individual study was interpreted as having a sufficient quality level when the average is above 2 points in each area.

For the inter-rater reliability of the qualitative analysis, a doctoral student majoring in special education who had understanding of QIs of Horner et al. (2005) participated in the analysis. Two from single-subject design studies were randomly selected. For those studies, we rated together on a 3-point Likert scale, ensuring that we had the same understanding of the concept and evaluation standard. After that, the author of the study conducted the analysis of QIs for 32 studies first. Then, the doctoral student reviewed the whole analysis results and identified the items that were rated differently. For the inconsistent items, we conducted an additional review and reached to the consensus through sufficient discussion. The reliability between two coders was 99.6%, which was calculated by dividing the number of consistent results by the sum of the number of both consistent and inconsistent results and multiplying it by 100.

2.3.2. Coding System for Meta-Analysis

A coding system for meta-analysis was developed by examining the prior coding systems designed for meta-analytic review on DBI (Filderman et al., 2018; Jung et al., 2018; Stecker, Fuchs, & Fuchs, 2005) as well as research on DBI concepts (NCII, 2013; Stecker, Fuchs, & Fuchs, 2005; Jung, Shin, & McMaster, 2016; Jung, 2018). The coding system for meta-analysis is presented in Table 1.

Table 1. Coding system for meta-analysis

Definition of Descriptors	
General Characteristics	
Year of publication	the publication year of study (2005~2020)
Type of publication	the type of publication, recorded as (1) journal article (2) dissertation
Participants	
Grade level	the grade level, recorded as (1) elementary school (2) middle school (3) high school
Sample size	the number of participants
Gender	the number of male and female participants
Exceptionality type	the disability or difficulties participants have, recorded as (1) with or at risk of learning disability (2) low achievement (3) under-achievement (4) intellectual disability (5) other
Intervention	
Topic of intervention	the main topic of intervention within academic domains (e.g., intervention on math using play activities)
Instructional approaches	the types of instructional approaches (e.g., direct instruction)
Setting	the type of instructional setting, recorded as (1) general classroom (2) special education classroom (3) special school (4) private center (e.g.,

	treatment center, hospital) (5) other (6) NA
Number of sessions	the total number of sessions
Length of each session	the length of each session (minutes)
Group size	the number of group members for instruction, recorded as (1) one-on-one (2) small (2-3) (3) medium (4-6) (4) NA
Interventionist	who provided the intervention, recorded as (1) researcher (2) teacher (3) undergraduate or graduate students (4) other (5) NA
Basic academic areas	the targeted basic academic domains (i.e., dependent variable), recorded as (1) reading (2) writing (3) math (4) other
DBI	
Type of CBM task	the type of CBM tasks (i.e., BASA) (e.g., reading fluency, early literacy, writing, early math, math)
Measurement frequency	the frequency of measurement with CBM for progress monitoring, recorded as (1) every session (2) every 2~3 sessions (3) every 4 sessions (4) other
Administrator of CBM	who administered CBM, recorded as (1) DBI instructor (2) other (3) NA
Data-based decision rule	the type of decision rule for instructional adaptation based on data from progress-monitoring, recorded as (1) slope-based rule (2) point-based rule (3) criterion mastery (4) other (5) NA
Instructional adaptation	the type of instructional adaptation, recorded as (1) quantitative changes (e.g., group size, time) (2) qualitative changes (e.g., instructional strategies, arrangement of environmental variables, type of feedback) (3) combination (4) NA

For the reliability of coding, the doctoral student who had research experiences using meta-analysis was participated in coding

for the meta-analysis as well. The author specifically explained the coding system and the coding method. Afterwards, three studies were randomly selected and coded by the two coders together. Then, the doctoral student reviewed the results that were firstly coded by the author, checking for the disagreed items. The two coders went through the process of additional review and sufficient discussion regarding the inconsistent items, resulting in consensus at the end. Through these procedures, the reliability of coding for meta-analysis was calculated as 98.86%.

2.4. Effect Size Calculations and Statistical Analyses

For 32 single-subject studies, the effect size calculation was conducted through the following procedures. First, the data points of each dependent variable were coded. When the study only contains the figures for outcome measures, a computer software program GetData Graph Digitizer (2013) was used to extract graphed data. Regarding the use of this computer program, it was reported to have high reliability and validity for extracting graphed data (Shin & Jung, 2018). Second, the standardized data points and the effect sizes were calculated using the method which had been suggested by Van den Noortgate & Onghena (2003a; 2003b; 2007) and used in Wang, Cui, & Parrila (2011), Wang, Parrila, & Cui (2013), and Heyvaert et al. (2012). To be specific, in order to transform raw scores into standardized scores, the raw scores of the data points were subtracted from the mean scores of the data points within the baseline and intervention phases and then divided by the standardized deviation of the data points from the phases combined (Wang, Cui, & Parrila, 2011). Then, the standardized mean score of the baseline phase was subtracted from each standardized score, which was to

adjust the mean of the baseline to 0. In this way, the standardized score of the intervention phase could be considered as the treatment effect. However, it should be noted that the standardized effect size calculated in this way cannot be directly compared to those calculated from group experimental design studies due to the difference in calculation method (Wang, Cui, & Parrila, 2011).

To synthesize the effect sizes calculated from single-subject studies, multi-level meta-analysis via hierarchical linear modeling was implemented. Meta-analysis has been increasingly being applied to single-subject design in order to empirically synthesize the results (Kratochwill & Levin, 2014). There is no consensus yet about what is the best statistical method or effect size indices when synthesizing the results from single-subject design studies (Gage & Lewis, 2014). However, it was pointed out that synthesis of results from single-subject studies tends to depend on visual analysis (Brossart et al., 2006; Van den Noorgate & Onghena, 2003a), which was probably subjective and difficult to compare the effectiveness across different studies (Wang, Cui & Parrila, 2011). On the other hand, meta-analysis using hierarchical linear modeling is another promising approach (Jenson et al., 2007; Van den Noorgate & Onghena, 2007). It can consider the issue of autocorrelation even when the number of measurements is relatively small and when the frequency of measurements is different across study participants (Van den Noorgate & Onghena, 2007; Raudenbush & Bryk, 2002).

Statistical analyses were computed with HLM software program version 6 (Raudenbush et al., 2004). To be specific, the overall effect size and the heterogeneity of effect sizes between study participants were examined. In single-subject studies, each study includes different study participants, and each participant includes time series

measurements, which represents the three-level model (Van den Noortgate et al., 2013). However, in the present study, the total number of single-subject studies was not enough to conduct a 3-level analysis. Thus, a 2-level meta-analysis was conducted. Level 1 represents the time series data points, and level 2 represents the participants of individual studies (see the following):

Level-1 Model: $Y_{ij} = \pi_{1j}(PHASE) + e_{ij}$ $\quad e_{ij} \sim N(0, \sigma^2)$

Level-2 Model: $\pi_{1j} = \beta_{10} + r_{1j}$ $\quad r_{1j} \sim N(0, \tau_{00}^2)$

The two levels represent the hierarchical structure of outcome variables (level 1) and study participants (level 2). Level-1 model shows a regression equation for the outcome variable. In the model, the outcome variable, Y_{ij}, represents the standardized score of the data point for occasion *i* and participant *j*; π_{1j} represents the effect size that equals the standardized difference of means between the intervention and baseline; *PHASE* is a dichotomous variable that represents the phase of each data point (i.e., 0 indicates that the data point is from the baseline phase and 1 indicates that the data point is from the intervention phase); e_{ij} reflects a random error term. Level-2 model suggests that the effect size of its corresponding participant, π_{1j}, equals the mean effect size across all participants (β_{10}) plus a residual (r_{1j}). Furthermore, an additional analysis was conducted corresponding to each basic academic domain of the outcome variables (i.e., reading, writing, and math).

After analyzing the mean effect sizes across study participants with the above unconditional model, several variables related to participants, intervention, and DBI were added to the model as

predictors in order to examine their impacts on the effectiveness. These predictors were added respectively, which was to avoid the potential risk of multicollinearity (Heyvaert et al., 2012). An instance of the conditional model is as the following.

Level-1 Model: $Y_{ij} = \pi_{1j}(PHASE) + e_{ij}$ $\quad e_{ij} \sim N(0, \sigma^2)$

Level-2 Model: $\pi_{1j} = \beta_{10} + \beta_{11}(GRADE) + r_{1j}$ $\quad r_{1j} \sim N(0, \tau_{00}^2)$

Additionally, there were several things considered in data analysis. First, when the intervention consisted of more than one phase, data from the very first intervention phase was coded to minimize the potential effect of confusing variables or time. In cases of the studies that implemented intervention in RTI approach, the first tier was coded as baseline, and the second or third tier intervention was coded as intervention phase for the analysis. Second, the scores on negative measures were reverse-coded. Third, multiple effect sizes from the same study or the same participant were avoided because of their dependence (Hedges, Tipton, & Johnson, 2010). Thus, when there were multiple measurements regarding the same construct, the representative one was adopted or the mean score was calculated when using the same scale. In contrast, when dependent variables were corresponding to different academic domains (e.g., reading fluency and computational skills), multiple effect sizes were calculated independently.

3. Results

3.1. Characteristics of the Studies

A total of 32 single-subject studies were selected for analysis. A total of 119 effect sizes from 1,601 data points and 100 participants were identified, which is summarized in Table 2.

Table 2. Number of participants and effect sizes

Study	Number of participants (m,f)	Number of effect sizes
Shim & Park (2015)	3 (0,3)	6
Shim & Park (2015)	3 (1,2)	3
Cho (2008)	3 (3,0)	6
Choi (2008)	1 (0,1)	1
You (2016)	3 (2,1)	3
Kim (2017)	3 (3,0)	3
Lee & Lee (2019)	3 (1,2)	3
Kwon & Kang (2007)	3 (3,0)	3
Yu (2010)	3	3
Han (2014)	3 (1,2)	3
Park & Lee (2015)	3 (1,2)	6
Lee (2015)	3 (3,0)	3
Kwon (2015)	3 (2,1)	3
Shin & Kang (2018)	3 (2,1)	9
Kim, Koh, & Yi (2014)	2 (1,1)	2
Kim (2011)	6	6
Ko & Song (2019)	1 (0,1)	1
Kim, Lee, & Lee (2014)	2 (1,1)	2
Lee (2015)	3 (3,0)	3
Kim & Lee (2015)	3 (2,1)	3
Gil (2015)	3 (3,0)	3
Seul (2005)	3 (2,1)	3
Kim et al. (2018)	5 (3,2)	5
Hwang (2012)	5 (2,3)	5
Ahn (2018)	4 (2,2)	8
Choi (2006)	4 (3,1)	4
Jung & Lee (2009)	4 (2,2)	4
Song & Kim (2020)	3 (1,2)	3
Kim (2015)	3 (2,1)	3
An (2012)	3 (3,0)	3

Jeong (2015)	3 (2,1)	3
Kim et al. (2020)	3 (2,1)	3
Total	**100**	**119**

Table 3 to Table 6 presents the characteristics of 32 single-subject studies in terms of general characteristics, participants-related characteristics, intervention-related characteristics, and DBI-related characteristics.

Table 3. General characteristics of the studies

Category		N	%
Year of publication			
	2005-2009	6	18.8
	2010-2014	8	25.0
	2015-2020	18	56.3
Type of publication			
	Journal	14	43.8
	Dissertation	18	56.3
Study design			
	Multiple baseline	12	37.5
	AB	5	15.6
	ABA	3	9.4
	ABAB	1	3.1
	Alternating treatment	1	3.1
	Other	10	31.3
Total		32	100

Table 4. Participant-related characteristics of the studies

Category		N	%	
Grade level				
Elementary school	Low (1-3)	19	59.4	87.6
	High (4-6)	7	21.9	
	Low + High	2	6.3	
Middle school		3	9.4	

High school	1	3.1
Exceptionality type		
With or at risk of learning disability	9	28.1
Low achievement	2	6.3
Under achievement	14	43.8
Intellectual disability	5	15.6
Other	2	6.3
School type		
General classroom	20	62.5
Special education classroom	6	18.8
Special school	1	3.1
Mixed	3	9.4
Other	2	6.3
Total	32	100

Table 5. Intervention-related characteristics of the studies

Category	N	%
Number of sessions		
10-19	23	71.9
20-29	7	21.9
Above 30	2	6.3
Number of sessions per week		
1	5	15.6
2	25	78.1
3	2	6.3
Length of each session		
15-30 mins	2	6.3
31-45 mins	23	71.9
46-60 mins	6	18.8
NA	1	3.1
Duration		
4-9 wks	22	68.8
10-14 wks	9	28.1
15-19 wks	1	3.1
Group size		
One-on-one	23	71.9
Small (2-3)	3	9.4
Medium (4-6)	2	6.3

Mixed (One-on-one + group)		2	6.3
NA		2	6.3
Interventionist			
Researcher		24	75.0
Graduate students		2	6.3
Researcher + Volunteer		1	3.1
Peer		1	3.1
NA		4	12.5
Instructional setting			
General classroom		16	50.0
Special education classroom		3	9.4
Special school		1	3.1
Private center		1	3.1
Other		5	15.6
NA		6	18.8
Basic academic domain (DV)			
Reading	Early literacy	1	3.1
	Reading fluency	7	21.9
	Fluency + comprehension	7	21.9
	Early literacy + Fluency + comprehension	1	3.1
Writing		6	18.8
Math	Early math	1	3.1
	Calculation	7	21.9
Reading + Writing		2	6.3
Total		32	100

Table 6. DBI-related characteristics of the studies

Category	N	%
Type of CBM task (BASA)		
Early literacy	1	3.1
Reading fluency	16	50.0
Fluency + Comprehension	1	3.1
Writing	6	18.8
Early math	1	3.1
Math	7	21.9
Measurement frequency (monitoring)		
Every session	20	62.5
Every 2-3 sessions	9	28.1

Every 4 sessions	2	6.3
Flexible	1	3.1
Administrator of CBM		
Interventionist	24	75.0
Other	2	6.3
NA	6	18.8
Data-based decision rules		
Slope	1	3.1
Mastery criterion	6	18.8
Slope + Data point	7	21.9
Slope + Mastery	1	3.1
Percentile	3	9.4
Other	1	3.1
NA	13	40.6
Instructional change		
Quantitative	1	3.1
Qualitative	6	18.8
Quantitative + Qualitative	3	9.4
NA	22	68.8
Total	32	100

3.2. Analysis of Quality Indicators

Table 7 presents the results of QI ratings for single-subject design studies. It was interpreted as having a sufficient quality level when the indicators were rated 2 or more. Rubric suggested by Chard et al. (2009) was referred for specific rating.

Table 7. QIs applied to single-subject design studies

Dimensions	Specific indicators	Number of studies			M
		1 (unsatisfied)	2 (partially satisfied)	3 (satisfied)	
Participants /setting	participants characteristics (e.g., age, gender, disability, diagnosis)	0	4	28	2.88

	process for selecting participants	1	5	26	2.78
	information about interventionists or teachers and comparability across conditions	21	7	4	1.47
	critical features of the physical setting	9	17	6	1.91
Dependent variable	description of DV	0	1	31	2.97
	measurement process	0	0	32	3.00
	measurement validity and description	0	4	28	2.88
	measurement frequency	0	3	29	2.91
	data collected on reliability (e.g., IOA=80%; Kappa=60%)	22	0	10	1.63
Independent variable	description of IV	0	8	24	2.75
	IV manipulation	14	7	11	1.91
	fidelity of implementation	22	1	9	1.59
Baseline	DV measurement	3	11	18	2.47
	description of baseline condition	12	16	4	1.75
Experimental control/Internal validity	design demonstrates experimental effect	0	0	32	3.00
	design controls for common threats to internal validity (e.g., elimination of rival hypotheses)	12	7	13	2.03
	patterns of results	0	0	32	3.00
External validity	replication of effects across participants, settings, or materials	2	11	19	2.53
Social validity	social importance of DV	0	0	32	3.00
	social importance of magnitude of change in DV	0	12	20	2.63

	practicality and cost effectiveness of implementation of IV	22	5	5	1.47
	typical nature of implementation of IV	16	15	1	1.53

3.3. Results from Two-Level Meta-Analysis

3.3.1. Results from the Unconditional Model: Mean Effect Size and Heterogeneity

The results from the unconditional model are displayed in Table 8. The mean of the effect sizes across all participants was 1.34. It was significantly different from zero, suggesting that DBI was effective in improving basic academic skills of participating school-aged students with learning difficulties. Additionally, the χ^2 statistic accompanying these variance components indicated that there was significant variability between the effect sizes among 100 participants in their effect sizes, suggesting the need to identify the possible predictors. Figure 2 visually shows 119 effect sizes from 100 participants and the variability between those effect sizes. In Figure 2, 0 and 1 on the x-axis represent the baseline and intervention phase respectively; the y-axis represents the standardized data points of outcome variables; and the slope of each line indicates the standardized effect size of each dependent variable. The figure also indicates that the effect sizes across participants were varying in a wide range.

Table 8. Results from the unconditional model

Fixed effect	Coefficient	SE	*t*	*df*	*p*-Value

Slope (β_{10})	1.341	0.078	17.224***	99	<0.001
Random effect	**SD**	**Variance component**	χ^2	***df***	***p*-Value**
Slope (r_{1j})	0.722	0.522	907.154***	99	<0.001
Level-1 error term (ϵ_{ij})	0.852	0.725			

**p<.05 **p<.01 ***p<.001*

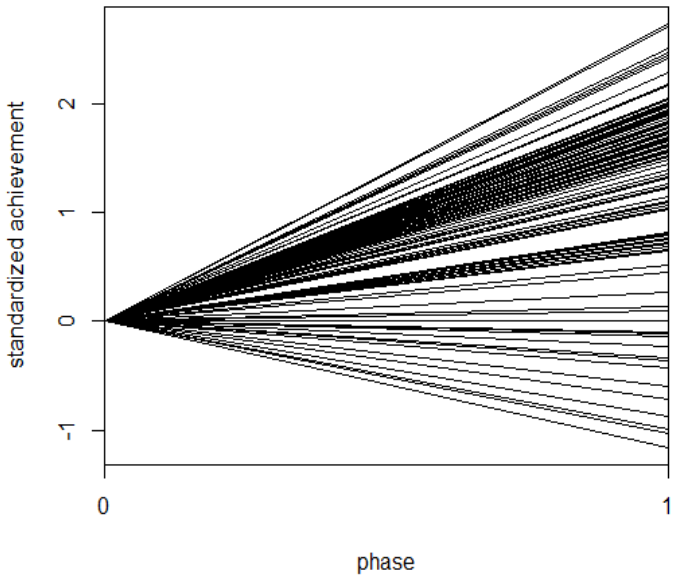


Figure 2. Effect sizes across participants

3.3.2. Mean Effect Size for Each Basic Academic Domain

Additionally, mean effect size for each basic academic domain (reading, writing, and math) was calculated. Phonological recognition, word recognition, reading fluency, and reading comprehension skills were included as outcome variables in reading. The studies included in writing were measuring the qualitative and quantitative scores, and

the studies included in math were assessing early mathematics skills (e.g., number recognition, counting) and computational skills. The results of the unconditional model for each area are shown in Table 9. For reading, 916 data points from 60 participants were included in the analysis, accounting for the largest proportion. 586 data points from 24 study participants were identified in writing, and 275 data points from 22 participants were identified in math. The mean effect sizes for domains of reading, writing, and math, were 1.432, 1.520, and 1.091 respectively, and they were all significantly different from zero. Furthermore, in writing and math, there were significant variability across participants in effect sizes.

Table 9. Results from the unconditional model for reading, writing, and math

Reading					
Fixed effect	Coefficient	SE	*t*	*df*	*p*-Value
Slope (β_{10})	1.432	0.090	15.939***	59	<0.001
Random effect	SD	Variance component	χ^2	*df*	*p*-Value
Slope (r_{1j})	0.633	0.400	429.965	59	>0.500
Level-1 error term (e_{ij})	0.833	0.694			
Writing					
Fixed effect	Coefficient	SE	*t*	*df*	*p*-Value
Slope (β_{10})	1.520	0.165	9.188***	23	<0.001
Random effect	SD	Variance component	χ^2	*df*	*p*-Value
Slope (r_{1j})	0.782	0.611	370.814***	23	<0.001
Level-1 error term (e_{ij})	0.802	0.643			
Math					
Fixed effect	Coefficient	SE	*t*	*df*	*p*-Value
Slope (β_{10})	1.091	0.188	5.815***	21	<0.001

Random effect	SD	Variance component	χ^2	*df*	*p*-Value
Slope (r_{1j})	0.823	0.678	208.508***	21	<0.001
Level-1 error term (e_{ij})	0.847	0.718			

*p<.05 **p<.01 ***p<.001

3.3.3. Results from the Conditional Model: Impact of the Predictors

Several variables related to participants, intervention, and DBI were added respectively to the two-level model as predictors in order to examine their impacts on the variability of effectiveness. Each variable was dichotomously coded for the analysis. For the missing data, the list-wise deletion was used. Table 10 shows the coding systems for the Level-2 predictors.

Table 10. Coding system for the Level-2 predictors

Predictors	Coding	
	0	1
Participant-related variables		
Grade level	Elementary school	Middle or high school
Exceptionality type	with or at risk of learning disability	Without or not at risk of learning disability
	With low or under-achievement	Without low or under-achievement
	With intellectual disability	Without intellectual disability
Gender	Male	Female
School type	General classroom	Other than general classroom
Intervention-related variables		
Number of sessions	19 sessions or less	More than 19 sessions
Number of sessions per week	1	2-3

Length of each session	45 minutes or less	More than 45 minutes
Duration	9 weeks or less	More than 9 weeks
Group size	One-on-one	Group
DBI-related variables		
Measurement frequency for monitoring	Every session	Every 2 or more sessions
Data-based decision rule	Including slope-based rule	Without including slope-based rule
Instructional adaptation	Specified	Not specified

Table 11 shows the results from the HLM model that included the characteristics related to participants as Level-2 predictors. The results indicated that the presence of intellectual disability significantly accounted for the variance in effect sizes (B = 0.532, p = 0.019). To be specific, the effect size was 1.304 for students with intellectual disabilities, whereas the effect size was 1.836 for those with learning difficulties without intellectual disabilities (i.e., with or at risk of learning disabilities, low or under-achievement, etc.). However, the effect sizes did not vary as a function of the other variables. It can be interpreted that the effectiveness of DBI was significant regardless of the grade level, classification of learning disabilities, classification of low or under-achievement, gender, and school type.

Table 11. Results of the conditional model with participant-related variables as predictors

Fixed effect		Coefficient	SE	t	df	p-Value
Grade level		0.043	0.161	0.267	79	0.790
Exceptionality	Learning disability	-0.237	0.201	-1.184	79	0.240
	Low or	-0.034	0.177	-0.195	79	0.846

type	under-achievement					
	Intellectual disability	0.532	0.222	2.394*	79	0.019
Gender		0.352	0.177	1.988	79	0.050
School type		0.312	0.184	1.694	79	0.094

*p<.05 **p<.01 ***p<.001

Table 12 shows the results from the model that included the intervention-related variables as Level-2 predictors. The number of sessions and the length of each session were the only predictors that significantly accounted for the variability in effect sizes across participants. Specifically, while the effect size was 1.552 when the intervention was provided less than 19 sessions, it was 0.984 when the intervention was continued for a longer period. In addition, whereas the effect size was 1.281 when the time per session was less than 45 minutes, it was 1.889 when each session was longer than 45 minutes. It was found that the effect sizes did not significantly vary as a function of the number of sessions per week, duration, and group size.

Table 12. Results of the conditional model with intervention-related variables as predictors

Fixed effect	Coefficient	SE	*t*	*df*	*p*-Value
Number of sessions	-0.568	0.189	-3.001**	79	0.004
Number of sessions per week	-0.222	0.225	-0.989	79	0.325
Length of each session	0.608	0.212	2.870**	79	0.005
Duration	-0.193	0.189	-1.019	79	0.311
Group size	0.006	0.217	0.026	79	0.979

*p<.05 **p<.01 ***p<.001

Table 13 presents the results from the model including DBI-related

variables as predictors. The coefficients of the variables were not significant, which indicated that the variation of the effect sizes among participants was not explained by the variables related to the features of DBI. It can be interpreted that the effect of intervention based on DBI was significant regardless of the specific variables related to DBI. However, it should be noted that the information about DBI process was not sufficient in many studies.

Table 13. Results of the conditional model with DBI-related variables as predictors

Fixed effect	Coefficient	SE	*t*	*df*	*p*-Value
Measurement frequency for monitoring	-0.164	0.185	-0.886	79	0.378
Data-based decision rule	-0.356	0.207	-1.714	79	0.090
Instructional adaptation	-0.301	0.194	-1.547	79	0.126

*$p<.05$ **$p<.01$ ***$p<.001$

4. Discussion

The purpose of this study was to identify the effectiveness of DBI for school-aged students with learning difficulties in Korea. Furthermore, the overall quality of the research that implemented data-based instruction for those students was determined in the study. The overall findings with respect to the research questions and directions for future research and practice were discussed below.

4.1. Research Trends on DBI

It was identified that research on DBI for students with learning difficulties have been conducted constantly in Korea since 2000s. It

indicates that there has been a continuous need to support diverse students with learning difficulties and ongoing efforts to apply DBI for those students. The review of the studies revealed that most of the studies were conducted for elementary school students in general classroom. Reading fluency was set as the outcome variable with the highest ratio. It is consistent with the previous findings that more than 80% of students with learning disabilities have difficulties in reading (Kim et al., 2017) and that reading fluency is a highly reliable indicator of the overall reading abilities, encompassing several linguistic skills in it (Abbott et al., 2012; Fuchs et al., 2001; Katzir et al., 2006). As for the characteristics of intervention, the results reveal that the most studies consisted of a total of 10 to 19 sessions, twice a week, and 31 to 45 minutes per session. Regarding characteristics of DBI, CBM for reading fluency was most frequently used, and students' progress was measured every session in the most studies. Additionally, the slope-based method and the point-based method were mostly used together, and instructional adaption was mostly conducted in qualitative aspects.

Despite these efforts to implement DBI in research and practice, it was identified that many studies did not report sufficient information about the principles or procedures of DBI in a systematic way. For instance, about half of the studies did not provide specific information on instructional change and data-based decision rules. Stecker, Fuchs, & Fuchs (2005) revealed that the frequency and quality of instructional changes have significant impact on the students' achievement. It not only benefits the students with learning difficulties but also teachers teaching them. For instance, in the study of McMaster et al. (2020), teachers who conducted DBI made instructional adaptions more frequently in more various aspects,

specifically based on convincing data rather than their intuition. Thus, future research and practice are needed to follow the procedures of DBI more systematically and to present the related information sufficiently.

4.2. Quality of Research on DBI

The overall quality of the studies was quite high. To be specific, 14 out of 22 Qis in single-subject design were rated as 2 points or more. Moreover, except for two, all studies scored an average of 2 points or higher. It indicates that research on DBI for school-aged student with learning difficulties have a sufficient quality in terms of research design and method. However, there were a few indicators scored below 2, which can be considered systematic weaknesses across studies. Specifically, in the future research, indicators regarding description of interventionist, comparability of interventionists across conditions, fidelity of intervention, manipulation of independent variable, and reliability of data collection are especially needed to be reported explicitly for the more reliable results.

4.3. Effectiveness of DBI

The effectiveness of DBI for school-aged students with learning difficulties was found to be quite high. The mean effect size from single-subject design studies was 1.34 and statistically significant. Although it cannot be directly compared with the effect sizes extracted from the meta-analysis of group design studies, it reveals that DBI is quite effective to improve basic academic skills of students with learning difficulties. It was suggested that students with learning difficulties tend to acquire and develop their learning skills at a slower pace than their peers (Bast & Reitsma, 1998;

Francis, 1992). It was also reported that for those students, intensifying intervention in a generic way is sometimes not enough to increase their achievement (Fuchs & Fuchs, 2015; NCII, 2013). Considering these previous findings, the result of the present study indicates that DBI is highly effective and suitable for those students. In addition, the use of CBM in DBI for assessing their achievement and growth seems to account for this large effectiveness. Shin et al. (2004) indicated that CBM tools can measure the ability and growth of students with learning difficulties in a more sensitive way. For instance, CBM in reading measures the reading abilities with texts that are equivalent and comparatively easy-to-read across different grades (Deno et al., 2001). Thus, the large effectiveness of DBI can be explained by the use of adequate measure for students with learning difficulties, and it should be noted that it is not only difficult but also unfair sometimes to check the progress of struggling learners with the identical measure used for those without learning difficulties.

4.4. Variability across Participants in Effect Sizes

The analysis of single-subject design studies revealed that there was significant variance across participants in effect sizes. The results from the conditional model showed that exceptionality type, the number of sessions, and the length of each session were significantly accountable for the variability of effect sizes.

Regarding the difference according to the presence of intellectual disabilities, the cognitive characteristics of students with intellectual disabilities seem to be associated to the comparatively lower effect size. In the present study, the number of participants with intellectual disabilities was 14, and the number of effect sizes was 17 in total.

The average of their IQ was 55.9, which is within the range of mild intellectual disability. The previous findings suggested that students with mild intellectual disabilities have educational needs distinguished from other students because of their cognitive features (Sabornie et al., 2005; Kirk et al., 2011). However, since the number of effect sizes was relatively small, it would require careful interpretation. Furthermore, it should be rather emphasized that the effectiveness of DBI for students with intellectual disabilities was significant as well and still was quite high (ES = 1.304). It means that DBI is quite effective and applicable for students with intellectual abilities as well as students without disabilities. Snyder & Ayres (2020) suggested that widely used measurement for basic academic skills such as the Woodcock Reading Mastery Test (WRMT; Woodcock, 2011) and Comprehensive Test of Phonological Processing (CTOPP-2; Wagner et al., 2013) are not suitable for those students due to its difficulty and unfamiliarity. In contrast, the general outcome measures such as CBM are regarded as more appropriate to evaluate the reading level and progress of students with intellectual disabilities (Wallace, Ticha, & Gustafson, 2008). Moreover, CBM is also recommended when making Individualized Education Plan (IEP) goals for students with intellectual disabilities in the areas of reading, writing, and mathematics (Mattatall, 2011). Thus, consistent with the previous findings, the results of the present study indicate that DBI and CBM can be applied for teaching students with intellectual disabilities.

Next, regarding the variance resulted from the number of sessions and the length of each session, more research is needed in the future. There are some inconsistent findings related to what is the most effective duration and length of sessions for DBI. For example, McMaster et al. (2020), where conducted DBI in writing for 20

weeks, recommended that teachers conduct interventions at least three times a week, 20~30 minutes per session. In the study of Bresina & McMaster (2020), the total number of weeks of DBI was found to significantly predict the improvement of students' progress. Therefore, further research on the optimal duration of intervention and time per session in DBI is needed.

4.5. Need of Teacher Support for DBI in Korea

When implementing DBI, the provision of support for teachers has been constantly suggested as the variable that significantly increases its effectiveness (Jung et al., 2018; McMaster et al., 2020). However, there were very few research that provided teacher support. In the study of Choi & Kwon (2009), it was revealed that many teachers in Korea were struggling with difficulties due to the lack of knowledge and information about instructional adaptations (Choi & Kwon, 2009). Therefore, it is required to provide understandable information and constant support for special and general teachers for implementing DBI. For instance, What Works Clearninghouse (WWC) platform from the U.S. Department of Education, which allows teachers to search, compare, and evaluate DBI for each basic academic area, and National Center on Intensive Intervention (NCII), which provides a tool chart and a checklist for monitoring progress and changing instruction, could be referred (Shin, 2019).

4.6. Study Limitations

Findings of the present study should be interpreted in light of the following limitations. First, although the studies which met the selection criteria were included in the analysis, sufficient information related to DBI such as data-based decision rules and instructional

changes were missing in many of them. Thus, there was a limit to analyze the impact of the probable predictors on variance in the effect sizes. Future study is needed to include more detailed information about DBI process, and there is need to analyze and compare the moderating effect of diverse variables. Second, two-level meta-analysis instead of three-level meta-analysis was implemented to synthesize the effect from single-subject design studies. It is recommended to conduct three-level meta-analysis of single-subject studies when there are more studies accumulated. Third, for calculating and synthesizing the effect sizes from single-subject design studies, the statistical method suggested by Van den Noortgate & Onghena (2003a; 2003b; 2007) and applied by Wang, Cui, & Parrila (2011), Wang, Parilla, & Cui (2013), and Heyvaert et al. (2012) was used. However, since there is no consensus yet for synthesizing the results of single-subject studies, it is necessary to calculate and compare the effect sizes calculated from various statistical methods. Fourth, in this study, only studies which used BASA as CBM tools were included for the analysis. It was because BASA was the standardized CBM widely used in basic academic areas in Korea. However, there is need that more various CBM tools with high validity and reliability should be developed and standardized in Korea. Then, future study is needed to synthesize the effectiveness and quality of DBI research using various CBM tools. Fifth, the mean effect size of DBI for each basic academic area was identified, but the difference in effectiveness depending on the specific focus of intervention or strategies was not analyzed in this study. Therefore, in the future study, the difference in effect sizes of DBI as function of specific instructional methods or contents should be examined. Sixth, in the present study, the QIs of Horner et al. (2005) were used

to evaluate the quality of individual studies. However, there is still need to examine the quality of the studies based on various criteria, especially beyond the aspects of research designs or methods. For example, Shin (2019) argued that the criteria suggested by Fuchs et al. (2017) and NCII, which reflect perspectives of teachers, should be considered when evaluating the quality of DBI.