Linköping Studies in Arts and Sciences, No. 946
Studies in Disability Research No. 121
Dissertation in Disability Research

Open Science for Evidence Synthesis

Improving research quality within intellectual disability special education research

Lucija Batinović

LINKÖPING UNIVERSITY

Linköping Studies in Arts and Sciences
Dissertations, No. 946
Studies in Disability Research No. 121

Open Science for Evidence Synthesis

Improving research quality within intellectual disability special education research

Lucija Batinović

Linköping University
Department of Behavioural Sciences and Learning
Disability Research Division
SE-581 83 Linköping, Sweden

Linköping 2026

At the Faculty of Arts and Sciences at Linköping University, research and doctoral studies are carried out within broad problem areas. Research is organized in interdisciplinary research environments and doctoral studies mainly in graduate schools. Jointly, they publish the series Linköping Studies in Arts and Sciences. This thesis comes from Disability Research Division at the Department of Behavioural Sciences and Learning.

CC BY

Typeset using LaTeX

Printed by LiU-Tryck, Linköping 2026

Edition 1:1

© Lucija Batinović, 2026
ISBN 978–91–8118–766–3 (print)
ISBN 978–91–8118–767–0 (PDF)
https://doi.org/10.3384/9789181187670
ISSN 2004-4887; 0282-9800 (print)
ISSN 2004-4895 (online)
Published articles have been reprinted with permission from the respective copyright holder. Figures presented in the thesis are created by Lucija Batinović, licensed under CC BY 4.0, unless stated otherwise.
Cover design: Lucija Batinović

POPULÄRVETENSKAPLIG SAMMANFATTNING

Grunden för varje policy som påverkar specialundervisning bygger på vetenskaplig forskning. Den vetenskapliga grunden för genomförandet av utbildningsinsatser säkerställer att insatserna är effektiva och inte skadliga för elever med intellektuell funktionsnedsättning. För att fastställa dessa insatser sammanställer forskare resultat från flera relevanta studier och drar allmänna slutsatser om effektiviteten genom en process som kallas evidenssyntes. Dock visar en växande mängd bevis att nuvarande forskningspraxis kan leda till mindre pålitliga resultat. Denna oroande trend kan påverka utbildningens kvalitet, och denna avhandling syftar till att förstå vilka forskningsmetoder som idag finns inom specialpedagogisk forskning för att utveckla metoder som förbättrar evidenssyntesen. Fokus i denna avhandling ligger särskilt på forskningsmetoder som främjar delning av forskningsmaterial och kunskap för att förbättra tillförlitligheten i resultaten. Den första studien identifierade specialpedagogiska insatser för elever med intellektuell funktionsnedsättning för att få en uppfattning om vanliga metoder inom detta område. Vi fann att det finns olika insatser som implementeras för att förbättra resultaten för elever med intellektuell funktionsnedsättning. Vidare lärde vi oss att dessa insatser generellt genomförs genom forskningsmetoder som fokuserar på ett litet antal individer, där insatserna genomförs under flera sessioner och resultaten mäts flera gånger före och efter att interventionen mottagits. Dessa metoder kallas single-case designs. Dessutom lärde vi oss att forskare sällan delar sina data och material, vilket gör det svårt att utvärdera studiernas trovärdighet och ytterligare förbättra insatserna. Den andra studien består av en evidenssyntes av single-case-designer för att ta reda på om läsinterventioner förbättrar ordigenkänningen. Baserat på de för närvarande tillgängliga resultaten från denna pågående studie fick vi veta att det är oklart hur effektiva phonics-interventioner är för att förbättra ordigenkänningen baserat på tillgängliga data. Den tredje studien beskriver skapandet av en checklista som forskare kan använda för att kontrollera om en studie är pålitlig innan den inkluderas i en evidenssyntes. Denna checklista hjälper forskare och beslutsfattare att fatta säkrare beslut om genomförandet av undersökta insatser. De för närvarande tillgängliga resultaten från en pågående studie visar att checklistan har potential att hjälpa forskare att identifiera problematiska studier och utesluta dem från syntesen. Den slutgiltiga artikeln introducerar en webbplats som innehåller vetenskapliga studier av utbildningsinterventioner. Denna webbplats gör det möjligt för forskare och praktiker att söka igenom studierna, läsa resultat, göra sin egen evidenssyntes och informera sina metoder med tillgänglig evidens. Dessa fyra studier vägledde en forskningsprocess som beskrivs i avhandlingen och som fokuserar på öppna vetenskapsmetoder vid evidenssyntes.

ABSTRACT

Evidence synthesis is the cornerstone of credible evidence-based practices, but the replication crisis and widespread documented questionable research practices have reduced the trustworthiness of evidence. The aim of this dissertation was to gain an understanding of the factors that affect the quality of evidence synthesis in special education research and develop methods to effectively synthesize available evidence, thus improving the quality of evidence synthesis. The dissertation is based on four studies of which three explored scientific practices in primary research involving school-based interventions for students with intellectual disability (Paper I), implemented new synthesis methods for prevalent research designs (Paper II), developed a research integrity evaluation checklist (Paper III). Namely, the findings indicate a prevalence of single-case designs and poor adherence to open science practices (Paper I). A preliminary synthesis of single-case designs illustrates how evidence synthesis can utilize shared data and facilitate further synthesis of research (Paper II). Preliminary findings also highlight the challenges in synthesizing underreported data, and indicate that contextual factors lead to heterogeneity. Preliminary findings from Paper III suggest usefulness of developed methods to reduce the impact of problematic studies in evidence synthesis. Paper IV introduces a platform for dynamic evidence synthesis of identified studies. The aim of the thesis is answered through the described open evidence synthesis workflow, guided by the findings of the four studies.

Contents

Abstract **iv**

Contents **v**

List of Figures **vii**

List of Tables **ix**

1 List of Papers **1**

2 Preamble **5**
2.1 General Aim 7

3 Background **9**
3.1 Disability Paradigms 9
3.2 History of Intellectual Disability within Special Education 12
3.3 Evidence-Based Practices in Special Education 13
3.4 Plurality of Perspectives in Special Education 13
3.5 Metascience in Special Education 16
3.6 Evidence Synthesis 22

4 Summary of the papers **27**
4.1 Paper I: School-based interventions for primary and secondary school students with intellectual disability: A scoping review 27
4.2 Paper II: Effectiveness of Reading Interventions on Word Recognition for Students with Intellectual Disability: A Systematic Review and IPD Meta-analysis of Single Case Experimental Designs 28
4.3 Paper III: Developing and Piloting a QUick Evaluation of Research Integrity for Evidence Synthesis in Education (QUERIES-Edu) 29
4.4 Paper IV: Living Review Platform for Educational Interventions: The Evidence in Learning and Didactics Community Augmented Meta-analysis 29

5 Open Evidence Synthesis Workflow **31**
5.1 Plan 31
5.2 Conduct 35
5.3 Disseminate 52
5.4 Supplement 54
5.5 Maintain 56

6 Implications **59**
6.1 Future directions 60

7 Conclusion **63**

Bibliography **65**

Acknowledgments **79**

List of Figures

3.1 Areas of metascientific research suggested by Ioannidis et al. (2015). 17
3.2 Open Science Research Workflow 20

5.1 Open Research Workflow for Evidence Synthesis 32
5.2 Individual trajectories of participants in reading intervention studies 42
5.3 Example forest plots with narrow and wide confidence intervals. 43
5.4 Average sample sizes per study design 44
5.5 Description of elements in a forest plot 46
5.6 Display of individual studies and systematic reviews on the platform 57
5.7 Example summary tab from the CAMA platform 58

List of Tables

5.1 QUERIES-Edu checklist items 39

1 List of Papers

This thesis is based on four papers, referred to in the text by their Roman numerals. The first paper is a pre-registered scoping review, published open access under a CC-by license. Paper II and Paper III are in-principle accepted (IPA) Stage 1 registered reports. Registered report format facilitates peer-review of protocols (Stage 1) before the study commences (Stage 2), and promotes acceptance based on study design, regardless of the outcomes, provided the authors adhere to their approved protocol and interpret the results in line with the evidence (i.e., in-principle acceptance). This format is a step towards reducing publication bias as it commits to publishing studies regardless of outcome, as long as there are no major deviations to the plan. These papers have been peer-reviewed by the Peer Community In Registered Reports. Peer Community In is a non-profit and non-commercial scientific organization that offers an open editorial process of scientific research. The entire peer-review process is openly shared alongside the edited paper. Both accepted protocols are available online as preprints under the CC-by 4.0 license. The thesis presents preliminary findings from the ongoing Stage 2 from both studies. The fourth paper is a methodological paper currently under review, and available publicly as a preprint under the CC-by 4.0 license.

I. Batinović, L., Topor, M., Henriksson, I., Allgulin, M., Pescador, X., Jonsson, E., Windsor, C., Krilanovic, A., Carlsson, R., & Danielsson, H. (2026). School-based interventions for primary and secondary school students with intellectual disability: A scoping review. *International Journal of Educational Research, 137*, 102960. https://doi.org/10.1016/j.ijer.2026.102960

II. Batinovic, L., Hammarsten, J., Nilsson, K., Topor, M., Holmer, E., Parlett-Pelleriti, C., & Danielsson, H. (2026). *Effectiveness of Reading Interventions on Word Recognition for Students with Intellectual Disability: A Systematic Review and IPD Meta-analysis of Single Case Experimental Designs.* PsyArXiv. https://doi.org/10.31234/osf.io/xu536 [Link leads to the IPA protocol, the preliminary Stage 2 is part of the dissertation]

III. Batinović, L., Topor, M., Hyltse, N., Danielsson, H., Brown, N. J. L., Cummins, J., Fillon, A., Heathers, J., Nilsonne, G., Westwood, S. J., Caquelin, L., Kalmendal, A., Nordström, T., Thoré, E.,

& Carlsson, R. (2025). *Developing and piloting a QUick Evaluation of Research Integrity for Evidence Synthesis in Educational Research (QUERIES-Edu).* In principle acceptance of Version 3 by Peer Community in Registered Reports. https://osf.io/tc5n3 [Link leads to the IPA protocol, the preliminary Stage 2 is part of the dissertation]

IV. Kalmendal Bruns, A., Batinovic, L., Claesson, F., Bylund, V., Blomqvist, A., Scholz, A. & Carlsson, R., (2026). *Living Review Platform for Educational Interventions: The Evidence in Learning Community Augmented Meta-analysis.* [Revised manuscript is part of the dissertation, preprint of an earlier version is available here: https://doi.org/10.31234/osf.io/hjku5_v1]

Glossary

AAIDD	The American Association on Intellectual and Developmental Disabilities
DSM	Diagnostic and Statistical Manual of Mental Disorders In this thesis, referring to the Fifth Edition, Text Revision (DSM-V-TR)
ICD	The International Classification of Diseases, in this thesis referring to the Eleventh Edition, ICD-11
ICF-CY	The International Classification of Functioning, Disability and Health for Children and Youth
APA	American Psychological Association
IDD	Intellectual developmental disorder. In this thesis, I also use the abbreviation ID to denote intellectual disability. In both cases, the thesis strictly refers to intellectual disability as defined by the DSM-V-TR, and does not refer to the broader category of "intellectual and developmental disabilities"
GRADE	Grading of Recommendations, Assessment, Development and Evaluation framework (Guyatt et al., 2008)
OSF	Open Science Framework
PCI-RR	Peer Community In Registered Reports. Peer Community In is a non-profit peer-review and dissemination platform for open science editorial practices. Peer Community In Registered Reports is a branch of the platform focused on publishing registered reports of studies, which includes peer-reviewing the protocol and the completed study within the same editorial workflow.
IPA-RR	In-Principle Acceptance Registered Report: a protocol accepted based on peer-review, after which the authors can conduct their study. The study is conditionally accepted, i.e., the completed study will be published given that the authors adhere to the accepted protocol and interpret their findings accordingly. Once those conditions are met, the study is published regardless of the outcome.
PICO(S)	Population, Intervention, Comparison, Outcome, and Study design
MECIR	The Methodological Expectations of Cochrane Intervention Reviews
MECCIR	The Methodological Expectations of Campbell Collaboration Intervention Reviews
JBI	Joanna-Briggs Institute
IPD	Individual participant data
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analyses; PRISMA-Scr is the extension for scoping reviews; PRISMA-IPD is an extension for IPD meta-analyses; PRISMA-P for protocols
SCED	Single-case experimental design; also abbreviated as SCD – single case design
SMD	Standardized mean difference
FAIR	Findable, Accessible, Interoperable, Reusable (M. D. Wilkinson et al., 2016)
CEC	Council for Exceptional Children
WWC	What Works Clearinghouse

2 Preamble

The experience of disability is inherent to human existence. This experience can be lifelong or short-lived, and can occur at any point in any person's lifetime. The Convention on the Rights of Persons with Disabilities adopted by the United Nations (UNESCO, 2006) defines persons with disabilities as follows:

> Persons with disabilities include those who have long-term physical, mental, intellectual or sensory impairments which in interaction with various barriers may hinder their full and effective participation in society on an equal basis with others.

This dissertation focuses on intellectual disabilities and methods with which science can aid in removing barriers to equal participation in society. Namely, the dissertation focuses on establishing high-quality evidence in special education research by implementing open science principles. Evidence-based special education is essential for improving academic achievement and long term quality of life for individuals with intellectual disabilities. Identifying which educational practices are effective requires rigorous and reliable synthesis of the available evidence. To this end, this thesis provides a comprehensive overview of the current landscape in disability research, focusing on special education interventions, to identify barriers and utilize open science practices to develop tools for reliable evidence synthesis, and ultimately reduce research waste. In this dissertation, I present where the research comes from, and then suggest where it should go.

Children with disabilities face challenges in education, signifying the importance of evidence-based practices (EBPs) to improve outcomes. According to the definitions from the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders, students with intellectual developmental disorder (IDD) have lower cognitive and adaptive skills (DSM-5-TR; American Psychiatric Association, 2022). This leads to challenges during education, and subsequently post-school outcomes. Bouck (2014) reports individuals with mild intellectual disabilities were found

to have lower employment rates, lower wages, and lower participation in postsecondary education compared to their peers without disabilities. Furthermore, another analysis of the National Longitudinal Transition Study-2 (NLTS2) data (Bouck, 2012) found that individuals with moderate or severe intellectual disabilities score low on post-school outcomes. However, research in special education often lacks adherence to open science practices and rigorous methods, even at the level of systematic reviews.

A way to ensure education leads to better outcomes is to implement evidence-based practices in education. While finding and conducting high quality research in education comes with challenges (e.g., difficulties conducting randomized controlled trials in schools), research in special education is even more complex, for its participants often vary in disability diagnoses (Odom et al., 2005). In their synthesis of what constitutes evidence-based practices (EBP) in special education, Odom et al. (2005) found that different professional institutions propose different constituents of evidence-based practices. For example, the American Psychological Association (APA) task force proposed frameworks for "well-established" versus "probably efficacious" interventions, while the Council for Exceptional Children (CEC) Division has not identified levels or types of evidence to guide practice. Finally, they mention the American Speech-Pathology Association framework which presents meta-analyses (including at least well randomized controlled studies) as the highest level of evidence for intervention effectiveness. In order to conduct a high-quality meta-analysis, a systematic review is necessary to ensure all relevant studies are discovered.

However, meta-reviews in education have revealed that even on the level of systematic reviews, following conduct guidelines and open science practices is still rare and poorly implemented (Nordström et al., 2023; Pellegrini et al., 2025). Research in education does not seem to be engaging in data sharing and preregistering methods, even when it comes to systematic reviews, which by definition have data publicly available, and do not experience privacy concerns. This puts an even higher strain on individual studies, where poor practices can be easily justified by rigorous rules and regulations aiming to protect the participants from privacy breaches.

This thesis presents the implementation of open science practices in evidence synthesis, guided by the findings from a comprehensive overview of educational interventions for students with intellectual disability (ID). As shown by Nordström et al. (2023) and Pellegrini et al. (2025), open science practices are still not part of the evidence synthesis field of educational sciences. Therefore, we planned to lead by example and conduct three evidence syntheses by following rigorous open science practices, preregistering our protocols and ensuring our results are trustworthy and transparent. Selecting areas where these syntheses were needed and feasible was not something we could improvise, so we planned to first scope all school interventions for students with ID to "get a feel for" the field, understand their methods and see what is prevalent. We wanted to use the resource to create a searchable database specifically curated for special education researchers.

The scoping review was pre-registered and conducted. The results however, led us towards a completely new direction. This is my acknowledgment of deviating from my initial plan for this thesis, which was to select three areas that are well-researched and suitable for a strong meta-analysis. We stumbled upon deeper issues with research integrity and open science practices, poorly followed reporting standards, and underdeveloped synthesis methods for research designs that were most prevalent. These findings ultimately shifted the aim from primarily synthesizing the existing evidence to improving the quality of research in the field.

Findings from four papers guide a proposed open evidence synthesis workflow, described in the dissertation, which encourages open science and research integrity practices at all following stages of the evidence synthesis process in special education: *scope the area, locate opportunities for meta-analytic research, critically evaluate the findings and contribute to the living evidence synthesis of the selected topic*. Finally, the dissertation discusses the implications of the findings to aid researchers and practitioners in understanding factors that affect the quality of evidence synthesis

in special education research, to facilitate development and implementation of evidence-based practices.

2.1 General Aim

The aim of this thesis was to gain an understanding of the factors that affect the quality of evidence synthesis in special education research and develop methods to effectively synthesize available evidence, thus improving the quality of evidence synthesis. Three research objectives were defined to achieve this aim, and were answered by the four studies presented in this thesis.

The first objective was to learn about the current state of research practices in the field of school-based interventions for students with intellectual disabilities. To answer this objective, I present results from Paper I, which was a comprehensive pre-registered mapping of school-based interventions for primary and secondary students with intellectual disability. The second objective was to determine methods to examine the effectiveness of educational interventions and facilitate open and reproducible evidence synthesis. To achieve this, Paper II was designed as a registered systematic review and a Bayesian meta-analysis. Paper II focused on developing a reproducible Bayesian and individual-participant data synthesis that utilized graphical data sharing practices in the identified single-case experimental designs. Additionally, the second objective was answered by the development of a framework for open and living evidence synthesis in learning and didactics. To support the use and systematic assessment of studies identified in Paper I, we developed an open infrastructure that allows users to upload new studies, and use the uploaded studies to conduct new evidence syntheses, and read or update existing ones. The platform is presented in Study IV and provides an interactive meta-analysis interface, and a searchable database of included studies and systematic reviews. The platform code is open source to encourage reuse in other fields. The final objective, which was to identify methods to ensure that evidence synthesis includes credible research practices to reduce the impact of research waste, was answered in Study III. Study III presents preliminary findings from a registered report modified Delphi panel and a pilot of a research integrity checklist. This checklist was developed based on the findings from Paper I, and aimed to expedite screening an overwhelming number of studies and ensure the studies included in the evidence synthesis are trustworthy.

Open Science Practices

Naturally, because this thesis focuses on research integrity and open science practices, here I acknowledge how I followed these principles in my own work.

Study I was publicly preregistered on the Open Science Framework (OSF) on June 16, 2022, before data collection started in August, 2022 (DOI: https://doi.org/10.17605/OSF.IO/R3ZHY). I preregistered the protocol together with the search strings and the data extraction template. During the process of conducting the review, there were deviations and updates to the protocol. These changes were transparently noted in the protocol that was uploaded to the accompanying OSF project, and all deviations were reported in the final manuscript in a separate section. The OSF project now also contains extracted data, both edited and "raw" extractions, the R script used for data cleaning and analysis, final conducted searches, screening procedure files, and original reference files from each database search.

Study II and Study III were conducted as Registered Reports. These protocols were submitted to Peer Community In Registered Reports (PCI RR) to receive an In-Principle Acceptance (IPA) before data collection. All accompanying materials, data, and scripts are available on OSF (Paper II: https://doi.org/10.17605/OSF.IO/8MXFH; Paper III: https://doi.org/10.17605/OSF.IO/NT548). Analysis files for Paper II are available on CodeOcean to allow users to reproduce the analysis without having a computer that

can run computationally heavy analyses. Study IV was not an empirical study, and therefore was not preregistered, but all accompanying materials are available on OSF (https://doi.org/10.17605/OSF.IO/QPMW5). Finally, all four papers are available as preprints, and Paper I was published open access.

Software I used when conducting the studies was generally free to use (e.g., Google Sheets, CodeOcean, Canva) and wherever possible, open source (e.g., R, R Studio, Zotero). The only proprietary, paid software were Covidence and Rayyan Pro, which was used to de-duplicate and screen studies in Paper I and Paper II.

Ethical Considerations

Unlike primary research, evidence synthesis relies on secondary data, which is generally made publicly available, and most often present in aggregated forms. For this reason, ethical approval is not needed to conduct evidence synthesis. The third paper, checklist development, involved data from two Delphi panels. The panelists however, are co-authors on the paper, and as collaborators agreed to provide data. Thus, the third paper also did not require ethical approval from The Swedish Ethical Review Authority.

However, ethical principles for conducting research synthesis have been followed, primarily through following rigorous conducting standards, minimizing researcher bias by preregistering data, sharing data and materials transparently, and facilitating reproducibility of all studies presented in the thesis.

Use of AI

AI (in this case, an umbrella term for machine learning, large language models or generative artificial intelligence) has been a great facilitator during my PhD. Although it is difficult to transparently lay out how I have used AI during my PhD, for the purposes of writing this dissertation, I have made a conscious decision to avoid using it to brainstorm, structure, or write the text. I have used AI (i.e., OpenAI ChatGPT Plus, Anthropic Claude, Google Gemini, Elsevier SCOPUS AI) to search for relevant literature, get example code for statistical figures presented in the thesis (Figures 5.5 and 5.3), or seek feedback for certain descriptions, without implementing direct suggestions for revision. Apart from the thesis, I have used AI to clean up code for my studies where necessary, to structure and format materials shared as supplements (e.g., write out code repository README files or create codebook templates, and instructions for screening or data extraction). Unlike for the thesis, I have used AI to review and suggest revisions for my submitted studies. AI is an excellent tool for academic research, and has helped me outsource tedious manual labour which allowed me to spend more time on learning how to do things properly. However, mental labour is mine, and should never be outsourced (no matter how much I would like it sometimes).

3 Background

3.1 Disability Paradigms

The focal constructs related to the concept of disability, defined by the World Health Organization (WHO) are disability, impairment, activity and participation. The WHO defines disability as "any condition of the body or mind (impairment) that makes it more difficult for the person with the condition to do certain activities (activity limitation) and interact with the world around them (participation restrictions)." Impairments encompass any significant deviations or loss in body structure or functions. Further, the WHO denotes activity as "an execution of a task", and participation as an "involvement in a life situations". There are several models proposed to explain the relationship between health, functioning, disability and impairments. In this section, I describe the most prominent models.

The long used medical model, which has historically been prevalent, describes disabilities as an "in body occurrence" and approaches disability from a biological standpoint. This means that the disability is viewed as caused entirely by an impairment, essentially equating impairments to disabilities. Under the medical paradigm of disability, which is the basis of the International Classification of Diseases (ICD), the experience of disability can be reduced by treating the impairment or providing adaptations to minimize the impairment effects (Goering, 2015). However, this impairment-centered approach left individuals with impairments feeling unaccepted by society, as they were pressured to comply with society norms. To counter the reductive paradigm of the medical model, a new model centered around external factors emerged. This social model of disability instead looked at how society can adapt to individuals with impairments. In turn, disability was viewed as a condition related to the external causes, e.g., participation restrictions (Goering, 2015). Nevertheless, the social model did not fully encompass the effects of impairment in experiencing disabilities, thus reducing the importance of medical interventions designed to treat impairments.

To address the challenges of the medical (also referred to as biomedical) model, new proposals of human functioning consisting of intertwined biological, psychological, and environmental components began developing in the past century. The pivotal moment of Engel's introduction of the biopsychosocial model of health (1977) caused a shift from approaching human functioning strictly from the medical (or physical) standpoint to seeing humans holistically, and has changed the way we conceptualize impairments and disability. Today, the International Classification of Functioning, Disability and Health (ICF-CY) understands disability as an interplay of health conditions and individual and societal factors, which corresponds to the framework of the biopsychosocial model.

The biopsychosocial model is a useful way of describing the relationship between internal and external factors of disability occurrence. There are different aspects of a person's functioning one can influence to eliminate negative effects of health conditions (i.e., reduce the experience of disability). What biopsychosocial model is able to explain is the interchangeability of different levels that are used to compensate when health conditions cannot be treated directly. For example, physical impairments of the eye (e.g., a cataract) can lead to the experience of visual disability (i.e., blindness). The experience of disability can be reduced by affecting the condition on the biological level, which simply means treating the impairment through the surgery of the eye. The treatment would enable the person to perform otherwise restricted activities, such as reading. Not all visual impairments can be treated, however. More importantly, not all individuals consent to the changes on the biological level. Nevertheless, a biopsychosocial framing allows the person to participate in activities by shifting to aids from the psychological and social levels when necessary. For example, visually impaired individuals can be accommodated by technological adaptations such as texts written in Braille or voice assistance. The BPSm sets reducing barriers to participation at the forefront, without focusing on resolving the underlying impairments. Thus, the BPSm is designed for society and the person with disabilities to meet halfway, and aid integration into society without reducing the independence of the person.

One of the factors that is not addressed by the BPSm is the effect of time on disability. Health conditions can change over time, individuals' experiences differ, and what might have been an applicable explanation of a disability at one point, might not be relevant in another. Therefore, it is necessary to continually evaluate which aspects of the BPSm are applicable in a given moment. Hypothetically, it might be beneficial to focus on providing accommodation on the individual level to each individual with intellectual developmental disorder during the initial diagnosis of the condition (e.g., early childhood intervention). However, the approach needs to evolve with time, as focusing on the psychological level throughout the progression of time could slow down improvements, which could for example be achieved by providing devices that could promote independence and allow participation. Thus, it is important to view the biopsychosocial model as a dynamic and constantly updating algorithm in order to avoid the use of the model becoming the cause of disability.

Intellectual Disability in the context of the BPSm

The Diagnostic and Statistical Manual (DSM-V-TR) defines intellectual disability (intellectual developmental disorder) based on deficits in intellectual functioning (measured by IQ tests), adaptive functioning, and onset in the developmental period (American Psychiatric Association, 2022). Similarly, according to The American Association on Intellectual and Developmental Disabilities (AAIDD), intellectual disability (ID) is a "condition characterized by significant limitations in both intellectual functioning and adaptive behavior that originates before the age of 22" (The AAIDD Ad Hoc Committee on Terminology and Classification, 2010). Global prevalence of intellectual disability is estimated generally around 1-3% (Patel et al., 2020), with substantial heterogeneity between high- and low-income countries (ranging from 0.05 to 1.55% in 2016, McKenzie et al., 2016). The Centers for Disease Control and Prevention (CDC) reported no significant increases in the incidence of intellectual disabilities in the United States between 2019 and 2021, estimated at 1.85% (Zablotsky et al., 2023). The prevalence has decreased throughout

the years, due to changes in the diagnostic criteria, and potentially due to prenatal screenings and pregnancy terminations. Aetiology of ID is not always known, and individuals with mild intellectual disability usually do not have a clear known cause. Known aetiologies are divided into environmental and genetic causes. Common genetic abnormalities that can cause intellectual disability are Down syndrome and Fragile X syndrome. Fetal alcohol syndrome is a common environmentally influenced cause of intellectual disability, which occurs due to maternal alcohol consumption during pregnancy (Denny et al., 2017).

Criteria for diagnosis is based on intellectual functioning (IQ below 70), poor adaptive behavior (i.e., "collection of conceptual, social, and practical skills that are learned and performed by people in their everyday lives"), and age of onset. Furthermore, intellectual disability is categorized based on severity of the disability according to the International Classification of Diseases (ICD-11) as: mild, moderate, severe, and profound (ICD-11 for Mortality and Morbidity Statistics, n.d.). Criteria for categorization puts more emphasis on adaptive skills and social functioning, and less emphasis on IQ test results, which implies the importance of participation restrictions when diagnosing intellectual disability.

Flexible application of the BPSm is particularly relevant for the development of educational interventions for students with ID. Intellectual disability can significantly impact an individual's ability to learn, communicate, and interact with others. Poor educational outcomes can result in unfavorable post-school outcomes. Bouck (2014) conducted a study analyzing data from the National Longitudinal Transition Study-2 which tracked the experiences of students transitioning from secondary education to adulthood. The study revealed that individuals with mild intellectual disabilities had lower employment rates, wages, and postsecondary education participation compared to their peers without disabilities. Additionally, Bouck (2012) found that individuals with moderate and severe intellectual disabilities showed poor performance on post-school outcomes. Following the principles of the BPSm, we can ensure that students are properly accommodated.

To overcome the barriers individuals with intellectual disabilities face, multiple acts were brought to guide the development of interventions that can help their integration in society. Although school interventions and curricula are not universally guided by the ICF model, there are countries that have implemented its framework into the special education laws (e.g., Portugal, Moretti et al., 2012). However, we can look at school interventions through the prisma of the BPSm.

For example, as suggested by the acts such as the UN Convention, multiple studies evaluated the effects of peer support on learning outcomes for students with intellectual disabilities (Biggs & Robison, 2023; Bowman-Perrott et al., 2023). Individual education plans (IEPs) have also been suggested as practices that could potentially improve educational outcomes for students with learning disabilities (McKenna et al., 2024; Southward & Kyzar, 2017). Peer support and IEPs can be seen as societal (environmental) adaptations that remove hurdles to educational participation for students with ID. Other school interventions have been focused on developing individual skills (psychological level) that would allow students with ID to overcome educational challenges. For example, there is a plethora of interventions that employ self-determination training to improve motivation and engagement of students with disabilities, to encourage them to complete education (Eisenman, 2007; Lindsay & Varahra, 2022; Shogren et al., 2012). Finally, there have been attempts to improve cognitive abilities through different interventions for example targeting working memory or reading (e.g., Danielsson et al., 2015; Ulriksen et al., 2024; van der Molen et al., 2010).

It is evident that there is a plethora of factors influencing the academic development of students with ID. Thus, it is important to be aware of the context surrounding students with ID to understand how to develop interventions which address all obstacles to participation and academic achievement.

3.2 History of Intellectual Disability within Special Education

Intellectual disability is associated with lower cognitive abilities that can disrupt learning. Danielsson et al. (2012) found that individuals with ID have poorer executive function skills than chronological age-matched typically developing individuals. Furthermore, they also perform worse on inhibition, planning, and non-verbal executive-loaded working memory tasks than mental age-matched typically developing individuals. Similar findings were reported by Costanzo et al. (2013) where individuals with Down and Williams syndrome performed worse on executive function tasks (e.g., planning, shifting, attention, short-term and working memory, etc.) than the mental age-matched typically developing group. However, areas with weaknesses differed between the individuals with Down syndrome (i.e., poorer performance in shifting and verbal aspects of memory and inhibition) and Williams syndrome (worse planning abilities). Systematic reviews by Spaniol and Danielsson (2022) and Tungate and Conners (2021) further corroborated these findings. These deficits have been linked to challenges with academic achievements, such as acquiring reading and spelling skills, and learning mathematics (Henry & Winfield, 2010; Tian & Liu, 2025).

Historically, intellectual disability was defined in various ways, and encompassed multiple impairments. Precisely, it was a part of disabilities under the "learning" or "mental" disabilities umbrella. Generally, students with ID were considered students with special education needs (SEN). Evidently, researchers and activists recognized the need to individualize and develop instructions and interventions to aid students with (intellectual) disabilities from various standpoints (when viewed from the BPS perspective). Special education needs have been perceived even earlier than the civil rights movements started, with the pioneer era in the United States going back to the late 1800s. As the disability acts emerged, researchers started recognizing the gap between the practice and existing evidence, which started the "research-based practices era" at the end of the 20th century (Collins et al., 2025). However, the perception of the public towards inclusion and educational betterment for students with disabilities only started changing in the 1960s, with the rise of the Civil Rights Movement in the United States and a similar movement emerging in the early 1970s in the United Kingdom (Baglieri et al., 2011).

The movements culminated at the beginning of the 21st century, with the establishment of several global acts that sought to ensure inclusive education for people with disabilities. The two acts covering multiple nations globally are the Salamanca Statement brought by UNESCO (1994), and the Convention on the rights of persons with disabilities by the United Nations (2006). According to the UN's Convention, the signed parties oblige to secure inclusive education for students with disabilities, through various means (e.g., "States Parties shall take appropriate measures, including: [f]acilitating the learning of Braille, alternative script, augmentative and alternative modes, means and formats of communication and orientation and mobility skills, and facilitating peer support and mentoring"). A similar proposal was established among European countries through the Equal opportunities for people with disabilities: a European action plan [2004-2010] (2003). Furthermore, a set of similar acts adopted in the United States were the Individuals with Disabilities Education Act (IDEA) from 1997 (and the Individuals With Disabilities Education Improvement Act (IDEIA) amendment from 2004) and the No Child Left Behind Act of 2001. For example, the IDEA is a set of laws that prescribe obligations from the States that ensure participation of students with disabilities in the educational process, sharing similar purpose to the acts from other nations. Essentially, students with disabilities should be provided with opportunities for post-secondary schooling or employment, adequate levels of achieved reading and writing skills, and access to qualified special education teachers and individualized academic curricula. These acts and conventions intend to determine societal support needed to remove obstacles to academic achievement and experience of disability by individuals with impairments, and serve as the basis for the development of teaching methods designed to accommodate students with disabilities.

3.3 Evidence-Based Practices in Special Education

The enactments adopted in the early 21st century emphasized the need for a systematic approach to evidence synthesis of research and its practical implications. These developments led to the introduction of evidence-based practices (EBPs) in special education by Odom et al. (2005, Collins et al., 2025).

According to Odom et al. (2005), evidence-based practices were adopted from the medical research field. In medicine, evidence-based practices were formed in the 1970s related to the movement intended to close the research-practitioner gap. General and special education soon after adopted the practice of research informing the teaching methods (Carnine, 1999). Although there have been numerous ways to describe bridging educational research and teaching practice, evidence-based practices could be considered the highest form of evidential value, when defined as "practices that are supported by multiple, high-quality studies that utilize research designs from which causality can be inferred and that demonstrate meaningful effects on student outcomes" (B. G. Cook and Cook, 2013, see also for other often interchangeably, albeit incorrectly, used terms like "research-based" practices). There are certain qualities that evidence-based practices should possess. Characteristics such as the research design of the assessed research, quality of the study, number of conducted studies, and the impact of the evaluated intervention or phenomenon (i.e., effect size) are assessed to conclude which practices can be labelled "evidence-based". The evaluated studies are generally located and synthesized through evidence synthesis procedures. Evidence-based practices are a crucial tool for providing safe and effective education to students with disabilities, and discourage the use of pseudo-scientific teaching practices.

To ensure EBPs are established in a standardized way, several frameworks have been designed, both in general and special education research. What Works Clearinghouse (WWC) is the most prominent organization, initiated by the U.S. Department of Education, that synthesizes educational research and provides evidence-based teaching recommendations. Similarly, the Council for Exceptional Children (CEC) focuses on special education, with the same goal regarding evidence-based practices. These organizations have developed strict guidelines for quality assessment of research reviewed to recommend evidence-based practices. Both WWC and CEC published methodological quality indicators for group and single-case experimental designs (Gersten et al., 2005; Horner et al., 2005; Kratochwill et al., 2012). The proposed guidelines are a noteworthy attempt to improve education, and despite inevitable limitations (see B. G. Cook et al., 2015; Maggin et al., 2022; Wolf, 2025, for examples), the guidelines are appropriately updated with new scientific discoveries and, conversely, the evidence base should follow the updated guidelines (cf. Williamson et al., 2019). Unfortunately, the translation of evidence to practice is still not universally prevalent, and teaching practices still commonly employ both evidence-based and non-empirically backed practices in their teaching (Burns & Ysseldyke, 2009; Carter et al., 2011; Hepburn et al., 2021). Furthermore, the lack of communication is bidirectional, as researchers do not seem to be guided by educators' needs nor evidence synthesis findings when conducting new studies (Day et al., 2025; Wald et al., 2024). Therefore, there is still a noticeable gap between education research and practices.

3.4 Plurality of Perspectives in Special Education

Baglieri et al. (2011) analyzed an ongoing division in the special education research between the incrementalist and reconceptualist researchers. The division was first clearly identified in 2000, where incrementalists were described as researchers that focus on influencing a change in the individual through educational interventions (cf. medical model), while reconceptualists believe a substantial shift toward accepting the individual and changing the environment (cf. social model). Furthermore, this divide trickles down to practitioners, as incrementalists approach research as a means of establishing evidence-based practices, while reconceptualists

call for changes in the way teachers' currently adopt practices without considering the effects on each individual with disabilities.

Baglieri and colleagues point to important issues raised by the reconceptualists that undermine the quality of traditionally conducted research. One of the examples was the inconsistency in diagnostic approaches and terminology. Specifically, the philosophical approach of conceptual relativism plays a critical role in this understanding. This perspective posits that our understanding of the world is mediated through language, conceptual frameworks, and various cultural or contextual agents that influence our perception of concepts, and reality in general. Taking the example of intellectual disability, while it is relatively concretely defined as a diagnosis in various medical and psychological manuals such as the DSM (Diagnostic and Statistical Manual of Mental Disorders) or the ICD (International Classification of Diseases), the operational definitions, nomenclature, and criteria vary considerably across different national contexts. For instance, what is termed as an intellectual disability in some countries is referred to as a learning disability in others. However, the term 'learning disability' may be defined differently in other countries, pointing towards a significant conceptual variation.

This diversity in definitions not only reflects the influence of different cultural and legal frameworks but also highlights the dynamic nature of medical and psychological terminology. Such variations can lead to discrepancies in diagnosis, treatment, and policy-making across different regions. By applying conceptual relativism, we can better understand and respect these differences, acknowledging that no single perspective holds an absolute view of truth but is rather shaped by the specific linguistic and cultural context in which it is embedded. On the other hand, it poses a threat to a higher quality of synthesis of work, as it is clear that there is poor consensus on what this area is conducting research on.

In terms of my ontological position and epistemological approach, I align myself with what Baglieri et al. (2011) describes as incrementalists. My research predominantly adopts a quantitative methodology, with a primary focus on tangible and measurable outcomes. While my research specialization primarily revolves around research methodology, particularly within the quantitative realm, issues raised from the reconceptualist side pose a serious threat to the validity and acceptance of traditional special education research and identification of evidence.

How we establish evidence

Within quantitative research, the gold standard for identifying evidence-based practices in the medical field has been the randomized controlled trial (RCT). RCTs belong to group designs, in which one group of randomly assigned participants receives an intervention, while another group of randomly assigned participants serves as control. Findings from these designs are then synthesized on a group level, representing the average effect of the intervention, putting less emphasis on individual variations between participants. Considering the diversity of the population with special education needs, it is natural that RCTs cannot always effectively establish relevant evidence-based practices in this area. The demographic heterogeneity and low incidence of conditions among this population resulted in a development of other methods which can more appropriately evaluate interventions, i.e., single-case experimental designs (SCED).

Single-case experimental designs are a common alternative to group designs in special education research. SCEDs are an adaptation of interrupted time-series experimental design, where researchers track a case over a longer period of time, with multiple measurement points, and analyze the effects of the disruption caused by introducing an intervention at a certain time point (Kratochwill et al., 2012). SCEDs have a multitude of variations for including the intervention phase, and are designed to establish a causal (functional) relationship between the independent variable (i.e., intervention) and the dependent variable within one case. Their specific appeal in special education is particularly the focus on within-case causality, and the possibility to establish this causality with just one case (hence the name single-case, or sometimes also referred

as N-of-1 trial). Special education research often involves individuals from low incidence populations, where large randomized-controlled trials would not be feasible. Furthermore, this population often comes with diverse needs and characteristics, which makes the sample heterogeneous and hard to meaningfully aggregate on population level.

Although SCEDs have been considered legitimate study designs for producing evidence-based practices, the design necessitates the same rigor for achieving internal, statistical, construct, and external validity, as do group (particularly RCTs) designs. Single-case experimental designs have evolved through the past decades and continuously adapt to applied setting needs and methodological challenges. A prime example is introducing randomization to the design, through multiple means. Among others, researchers can randomize the order of phases within participants or the order of intervention introduction between participants (Kazdin, 2021). There are, however, genuine causes for concern in the SCED research area, which can be separated into two categories of issues: those inherent to the design, and those related to the researcher practices.

The latter involves over-reliance on visual analysis, and high publication bias. Traditionally, the effect (i.e., functional relationship) of an intervention in SCEDs is inferred through visual inspection of the graphed data from each session. This process usually involves identifying the level of variability between sessions, and the difference between the intercept and slope of the intervention phase compared to the baseline phase. Originally, visual analysis allowed researchers to identify only clear, strong effects, but as Kazdin (2021) points out, this risks overlooking important, but weaker effects, and more importantly, risks false positive findings. Second, as the field has not been highly regulated when it comes to reporting and transparency standards, there is a tendency to publish only significant results, and researchers even omit results from participants with non-significant results (Shadish et al., 2016). Of course, these problematic practices are far from exclusive to SCED research, and the issues can simply be resolved by just not doing them.

More complex are the issues related to the nature of the design. Petursdottir and Carr (2018) provide a taxonomy of validity threats to SCEDs, specifically in Applied Behavior Analysis. Petursdottir and Carr laid out validity threats for the four categories of validity (statistical, internal, construct, and external) and divided them into those with higher and lower relevance to SCEDs, compared to group designs. Of particular threat to educational interventions, which cause irreversible effects on the participant, are history, maturation, and testing. History refers to any incident occurring during the experiment that is taking place, which can affect the implementation and effects of the intervention. For example, the intervention start can coincide with the student being transferred to a new class, which could disrupt their learning abilities. Generally, history can be an incident that only applies to one of the participants, in which case that effect can be assessed by comparing to other students. However, an incident affecting every case in the experiment might pose a more serious threat. Maturation equally affects the state of the participant, but is internally driven, and relates to physical, psychological and physiological changes that just occur naturally with time progression. For example, a child's cognitive development happens naturally through time, even when there is no intervention implemented. A child will naturally pick up language as time passes, regardless of a language intervention being implemented. Maturation effects can therefore obscure intervention effects. Slocum et al. (2022), however, summarize different arguments on whether these validity threats are equally present in different SCEDs. Specifically, one argument is that maturation as a process is slow and gradual, and for most interventions which last weeks or months should not have an impact. Although these arguments are primarily focused on multiple-baseline design, this reasoning can apply to other SCEDs. Testing is a methodological threat that stems from repeated implementation of the same assessment, which is at the core of SCEDs. When participants get repeatedly tested on the same topic (often using the same measurement), there is a risk that the learning might occur just from frequent repetition of the task.

Although these threats are not exclusive to single-case experimental designs, RCTs handle the implications better through randomization which ensures these effects are equally spread across

all participants, which minimizes their impact. Particularly, chastising single-case designs for the lack of rigour in conducted studies implies that existing RCTs in the field are universally conducted properly and emulate the gold standard of the design, when the latter is far from the truth. Taken together, although possible disadvantages, SCEDs offer a new and needed approaches to address educational needs of students with intellectual disability.

Plurality of perspectives is a necessary precursor to the development of accommodations for a population such as intellectually disabled students, which is rare in incidence and complex in the variety and severity of co-occurring conditions. Acknowledging methodological weaknesses of established interventional approaches is crucial for maintaining an open mind towards advancing evidence-based practices. To accurately address the pitfalls and establish high-quality methods for improvement of special education, one solution is to approach this dilemma from a meta-scientific perspective.

3.5 Metascience in Special Education

Metascience, otherwise known as meta-research or "research on research", is a scientific field that has a long, multidisciplinary history, drawing on statistics, philosophy, sociology, scientometrics, and research synthesis. Although it emerged as a more unified field in the wake of the replication crisis in early 2010s, the field has been developed through the works in the area of philosophy of science, with the contributions of several notable researchers throughout history, such as Paul Meehl and Imre Lakatos (Faust, 2005). Essentially, metascience uses approaches from philosophy of science to evaluate how research is done, and how it can be improved. Among his works, Meehl has focused on meta-theories (theories on theories; Meehl, 1967) and weaknesses within research syntheses that affect evidence-based practices (Meehl, 1990)). Although Meehl was a psychologist and based most of his work in metascientific developments within psychological fields, metascientific work has influenced major changes in the medical field, beginning with a seminal paper assessing statistical methods of over 200 medical publications (Schor & Karten, 1966), whose findings led to questioning the research validity and replicability in medical fields, and that doubt spilled over to other soft sciences (particularly psychology).

As Lakens (2025a, 2025b) discussed, social psychology has been going through a confidence crisis for decades, with the first well-known research crisis appearing in the 1960s. The first crisis indicated that the research output proliferated, but the substantive content was rare, and most of the research was of poor quality. Issues that emerged as critical were tied to undefined ethical standards and methodological issues regarding sample quality and experimenter effects. These problems were mostly addressed through the development of ethical review boards and improvements in the methodological standards. However, replication concerns existed already then, and have lingered on until today.

Currently, the dominant paradigm in educational quantitative research focuses heavily on producing interesting, novel, and often counterintuitive results, which leads to high citation counts and low p-values. “Features” of research like low p-values and being highly cited have become indicators of quality and rigor. Under this paradigm, the empirical methods endorsed by positivism were highly valued, and the integrity of research was often evaluated by these metrics. However, the second replication crisis—an event born from questioning the reliability and validity of many research findings in social psychology—introduced a shift towards a new paradigm (see Lakens, 2025a, 2025b, for an overview).

Replication issues culminated with the second crisis, which has been ongoing for the past two decades, and has poured over to other research disciplines. The second replication crisis was soft-launched with the seminal paper by John Ioannidis, published in 2005, “Why most published research findings are false”. In this paper, Ioannidis reflected on the research practices in medical research (of which many were an issue already in the first confidence crisis) which

Figure 3.1: Areas of metascientific research suggested by Ioannidis et al. (2015).

Meta-research area	Specific interests (nonexhaustive list)
Methods: "performing research"—study design, methods, statistics, research synthesis, collaboration, and ethics	Biases and questionable practices in conducting research, methods to reduce such biases, meta-analysis, research synthesis, integration of evidence, crossdesign synthesis, collaborative team science and consortia, research integrity and ethics
Reporting: "communicating research"—reporting standards, study registration, disclosing conflicts of interest, information to patients, public, and policy-makers	Biases and questionable practices in reporting, explaining, disseminating and popularizing research, conflicts of interest disclosure and management, study registration and other bias-prevention measures, and methods to monitor and reduce such issues
Reproducibility: "verifying research"—sharing data and methods, repeatability, replicability, reproducibility, and self-correction	Obstacles to sharing data and methods, replication studies, replicability and reproducibility of published research, methods to improve them, effectiveness of correction and self-correction of the literature, and methods to improve them
Evaluation: "evaluating research"—prepublication peer review, postpublication peer review, research funding criteria, and other means of evaluating scientific quality	Effectiveness, costs, and benefits of old and new approaches to peer review and other science assessment methods, and methods to improve them
Incentives: "rewarding research": promotion criteria, rewards, and penalties in research evaluation for individuals, teams, and institutions	Accuracy, effectiveness, costs, and benefits of old and new approaches to ranking and evaluating the performance, quality, value of research, individuals, teams, and institutions

doi:10.1371/journal.pbio.1002264.t001

Note. Reproduced from Ioannidis et al. (2015), licensed under CC-BY 4.0.

lead to overstated conclusions that are not substantiated, nor repeatable. Factors like research bias, small sample sizes and flexibility in decisions regarding the design degrade confidence in the findings and often obfuscate reality. Following this paper, the methodological crisis started spreading, and could be considered hard-launched with two critical papers which empirically confirmed Ioannidis' insinuations. The first one was published by Simmons et al. (2011), and introduced the field to the concept of questionable research practices. The second paper was a large-scale attempt to replicate a 100 reports, which was led by the Open Science Collaboration and published in 2015. Out of the 100 replicated studies, only 36% of the significant findings were replicated. The second crisis led to the establishment of metascience as a new research field. Metascience is a scientific field that researches scientific practices, methodology, production of research and its dissemination with the intent of understanding and improving scientific procedures. Although not new, the metascientific discipline has recently started evolving into a structured research area with the generation of new methods, research practices, and increased contributions from the research community. Some of the scientific areas that metascience researchers cover are philosophy of science, evidence synthesis, sciento- and bibliometrics, reproducibility, and evidence-base practices (Ioannidis et al., 2015, see reproduced Figure 3.1).

Although there are many meta-research areas, two closely related areas of meta-science within reproducibility and methods are the focal points in this thesis: open science and research synthesis. Both areas are crucial to identify trustworthy evidence-based practices in special education. These meta-scientific approaches stemmed from a paradigm shift. Thomas Kuhn's scientific paradigm describes progression and conduct of science within the context of current norms and values specific to a certain time period (Benton & Craib, 2011). A paradigm shift thus occurs once the anomalies of the current paradigm build into a crisis. According to Kuhn, scientific

research is deeply influenced by the dominant paradigms, which dictate what questions are asked, how investigations are structured, and how results are interpreted. This framework is particularly illuminating when applied to the field of psychological research, which has experienced significant shifts in paradigms, especially noticeable before and after the replication crisis. Although this shift is yet to reach educational research, the meta-scientific perspective I am taking is heavily influenced by the newly established paradigm. This newer paradigm emphasizes transparency, reproducibility, and the incorporation of open science practices, such as pre-registration and sharing data for reanalysis, and it birthed the modern metascientific approach to research.

The transition between paradigms has led to a clash within the existing educational research and my position as a reviewer of this literature. Older studies that dominate the field based on traditional metrics of success are now scrutinized under the new standards, which prioritize different criteria. The lack of open science practices, which used to be overlooked, now jeopardizes the positive evaluation of these studies, despite their potential genuine quality. As newer researchers begin to adopt and advocate for this current paradigm, there will be a growing divide between what was once considered rigorous and what is now deemed essential for ensuring the validity and reliability of research outcomes. This shift makes it difficult to assess the state of the literature, as the two realities exist parallelly. On the one hand, assessing the literature based on the standards during their times, and ignoring what we know now, appears obtuse and irrelevant. On the other hand, assessing the literature according to today's standards could be argued equally pointless, as they are deemed to fail the evaluation process (cf. Williamson et al., 2019). Regardless, improvement of the field lies in developing techniques that increase trustworthiness of future research, even if we risk losing confidence in the past research.

Open Science and Reproducibility

The establishment of a more unified field of metascience resulted in changes in how we approach dissemination of the entire research cycle. One of the metascientific answers to the problem of verification and replication was the rise of open science (Nosek et al., 2015; Nosek et al., 2012). This term reflects a scientific culture of approaching science with transparency in one's work, openness to external verification and care for robustness over novelty. While open science is an umbrella term, one attempt to define it is an initiative for "transparent and accessible knowledge that is shared and developed through collaborative networks" (Vicente-Saez & Martinez-Fuentes, 2018). The scientific endeavors of open science enthusiasts have radicalized the approach to conducting research, and, although still not accepted universally, make strides towards implementing these practices in research infrastructure worldwide.

Practices that qualify as open science have existed for centuries, and take various forms. In the modern context, the works of librarians and information specialists to provide open access to knowledge is the most noteworthy example. One of the early catalysts of open research was the development of a preprint repository `ArXiv` launched in 1991. Although early proponents of open access set the ground for the development of open science standards, the movement evolved to today's standards with the efforts in the early 2010s, led by Simine Vazire, Brian Nosek, Tal Yarkoni, Chris Chambers, and many others, which have resulted in infrastructure and frameworks that encourage and facilitate adhering to open science practices. For example, one of the largest organizations for open science endeavors, the Center for Open Science was founded in 2013 (unofficially by launching the Open Science Framework platform in 2012). Closely related, the Society for Improvement of Psychological Science (SIPS) was launched in 2016 in response to the replication crisis. Another effort to improve science within the open science culture was the introduction of a registered reports format (Chambers & Tzavella, 2022). Locally (but still internationally), a Swedish community-run journal *Meta-Psychology* has been one of the pioneers of a completely open peer review process with stringent requirements for open science adherence since 2017 (Carlsson et al., 2017). Particularly important work comes from a renowned neurodevelopmental disorders researcher, Dorothy Bishop.

Bishop contributed to adoption of open science and reproducibility practices, and contributed to improving the state of research through meticulous error assessments of numerous studies in disability research. While these were initiatives from passionate individuals, pushed forth by grassroot-community efforts, they culminated with major international organizations adopting open science principles. Today, 194 countries follow UNESCO's 2021 recommendations to implement open science practices and promote transparency, collaboration and citizen science to improve rigor and confidence in research (UNESCO, 2021). This means that researchers should follow FAIR principles (Findable, Accessible, Interoperable, and Reusable; M. D. Wilkinson et al., 2016) when sharing their data, analysis code and materials.

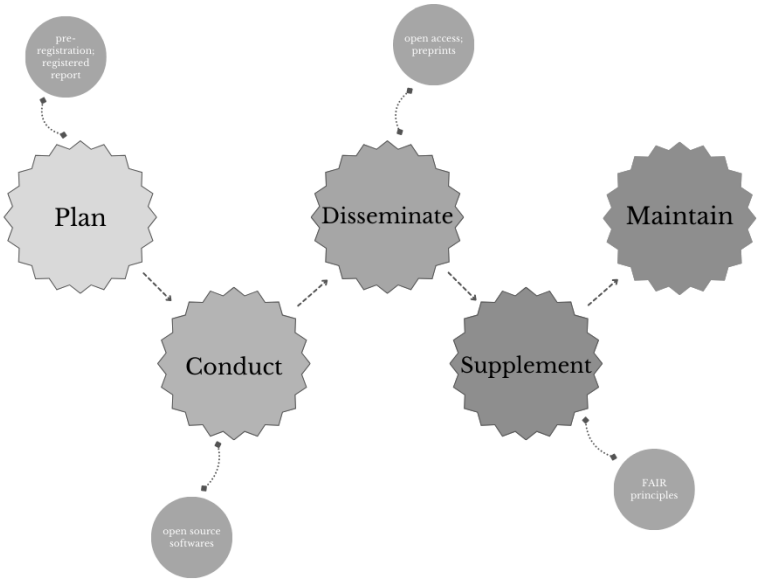
Appeals for inclusion of open science practices are not foreign to special education research. B. G. Cook et al. (2022) published a guide for open science practices in special education. However, the practices have ostensibly not yet been picked up, as the review of open science practices shows little growth over the years (B. G. Cook et al., 2025; Eiser Hess et al., 2025; Makel et al., 2024). Familiarity or low coverage does not seem to be an issue, as Makel et al. (2021) found that special education researchers report adhering to open science practices as well as to questionable research practices, and Fleming et al. (2025) found similar reports on attitudes toward open science practices. Although they self-report adherence, Cook et al. (2023; 2025) findings do not corroborate those claims. More attempts from Cook to introduce open science practices were proposed over the years. Cook et al. (2025) recently introduced registered reports in learning disability research and calls for greater adoption of open science practices in the special issue examining the state of evidence in special education research. There is a clear divide between a group of enthusiasts in the field and the current state of requirements from the publishing venues and general research community, mostly due to poor accessibility and education about integrating open science in regular research. The slow change towards more robust research practices thus affects the quality of evidence upon which practices are based.

Open Science Research Workflow

Based on the developments in open science, we can improve transparency practices and foundations for quality evidence. Figure 3.2 depicts a modified workflow diagram for a research study that follows open science practices from its inception and continues beyond publication. Different open research cycle diagrams exist (e.g., French, 2022), but in my view, they do not completely represent the workflow stages, and for my purposes, they should not be circular. The workflow encourages incrementality of research by defining a "maintenance" stage at the end of the workflow, which is supposed to continually be updated with new findings. This continuation of evidence can be exemplified through living reviews of existing evidence, or multilab studies within primary research. Further, the maintenance stage is meant to elucidate research gaps through the accumulation of evidence. The new research resulting from that stage can be seen in a circle form, as the new research should follow the same stages of the workflow. Nevertheless, that is beyond the scope of this dissertation. Chapter 5 describes the process of conducting evidence synthesis within the open research workflow, based on the studies included in this dissertation.

The workflow presented in Figure 3.2 consists of a study planning stage, followed by the conduct, dissemination, supplementing and maintenance. The first two stages represent the bulk of the work, where the planning stage should aspire to reduce the amount of work done during the following stages of the workflow, up to and including maintaining. Each study should start with a study design and plan. This is a crucial step, where we can mostly affect how trustworthy our results will be. As Simmons et al. (2011) and Ioannidis (2005) pointed out, HARK-ing (hypothesizing after results are known) and questionable research practices (QRPs) are detrimental for replicability of results. Therefore, transparent and detailed planning which includes stating in advance what we expect to find, and how exactly we will look for evidence, allows readers to evaluate the severity of tests and trustworthiness of reported findings (Lakens, 2022; Lakens et al., 2024).

Figure 3.2: Open Science Research Workflow



Note. The elements of this diagram are loosely based on the open science cycle steps (e.g., French, 2022).

Throughout the years, multiple platforms and templates have been developed to facilitate sharing study protocols openly, i.e., pre-registering research. Platforms like the Open Science Framework (for any type of research) or PROSPERO (for medical research evidence synthesis) provide templates and infrastructure for researchers to clearly lay out their hypotheses, analysis plan, sampling justification, and supplement any materials and code needed to understand and verify the planned study. Fleming et al. (2023) provided a review and a guide on different pre-registration platforms suitable for special education research. One step further is to write a registered report. A registered report is essentially a pre-registration, with the benefit of receiving a peer review on the planned study, before it is conducted, but unlike a preregistration published in a journal, scientists are "rewarded" with an in-principle acceptance (IPA), which means that their study will be published even if the results are not significant or do not support the hypothesis. This relieves the researchers of the need to provide significant and novel findings to publish their work. As long as they follow the planned protocol, the results will not affect the publication opportunities. The Cochrane and Campbell Collaboration have long implemented this system for systematic reviews they publish, and since recently, more journals are beginning to adopt this practice for other types of empirical research.

Once we have transparently committed to a study plan, we can conduct the study. While there is less emphasis on open science practices at this stage of research, there are many opportunities to follow open science principles during data collection and analysis. Prioritizing open source softwares for data collection and analysis promotes accessible research that can more easily be reused by other researchers, and allows collaboration between researchers with various resource limitations. For example, conducting the analyses in R or Python allows collaboration between multiple collaborators without worrying about licensing availability or funding constraints.

After the data is collected, analyzed, and the article is written, we should consider how to disseminate the research. Open access has been one open science practice that was implemented for the longest time, and is most likely the most widespread convention of all open science practices. It is natural that open science starts with open access to the scientific report. All supplemental materials will rarely have the reach and impact as a report of a conducted study would, as the latter is what we build knowledge on. Open access articles reach wider audiences, and are often more cited (Huang et al., 2024). Accessibility to research is crucial for the incremental nature of science, and promotes diversity and equality. Therefore, studies should be published open access whenever possible. Journals provide various open access alternatives, and diamond open access journals are the most favorable ones, as they do not charge authors to publish their work open access. Alternatives to journals exist, and although they are not all peer-reviewed, preprint platforms allow authors to share their work early and open access, often while they wait for peer-review. Pre-prints are academic papers which are published online but have not been peer-reviewed, on platforms like `PsyArXiv` or `MetArXiv`. These platforms provide moderation, mostly on technical level, to ensure the scope fits their platform and that all basic features of the work are present (e.g., the report contains proper metadata, and the basic content like methods or results section). Pre-prints are beneficial as they reach the audience quickly and allow incorporating broader feedback to their work while providing version control and digital object identifiers (DOI), which is a great way to break barriers set by the traditional publishing system which is often costly and gatekeeps research from poorer research communities.

While dissemination of research reports is the foundation, supplemental materials are what discerns unverifiable and verifiable research. The major improvement caused by the second replication crisis came from insisting on sharing underlying data, code and materials needed to reproduce and replicate findings. Thus, each published study should supplement the findings following the FAIR principles. This means that the data, code, and materials should first be made available on platforms that are preferably open, findable through simple search means, with persistent links attached to them. For example, the Open Science Framework or Zenodo offer infrastructure to share datasets, registrations or preprints for free, and are accessible to anyone external from the researcher. Accessible files are files that can be opened and used as they are, available to anyone with accesto a computer and internet. However, just finding and accessing data or code is rarely enough to verify results or replicate findings. It is often difficult to understand our own data and code after some time away, let alone for a complete outsider. Hence, shared supplements need to be interoperable and reusable as well. Interoperability of files concerns the structure of the dataset and attached metadata that allows different algorithms to parse them in the same way. Interoperability facilitates updating and merging to other data points. Last, but not least, shared materials should be reusable. Reusability is perhaps the most complex component, as it requires time and effort to provide materials that are coherent to other users without consultation with the original authors. Today, LLMs can help with the manual labour of annotating code, explaining variables, and providing detailed metadata, but it is still crucial to deeply think about the structure and elaboration accompanying supplemental materials to increase reuse and expedite verification checks.

The final step of the open science workflow is not considered or emphasised often, as a study is often considered concluded as soon as it is published. Despite the conventions, open science enthusiasts should maintain their research in whichever way it is necessary. This could involve being responsive to enquiries from other researchers about asking for access to or explanations of the data. Similarly, it involves being open to corrections and critiques from other authors, either through post-publication review platforms (e.g., PubPeer) or privately. For research syntheses, it can also include maintaining and timely updating the evidence base. These practices are dismissed by the traditional venues, but are a basic requirement for conducting responsible science.

Improving Research Quality and Reducing Research Waste

The proliferation of diverse research, especially driven by the plurality of perspectives and approaches to tackling special education needs, is inevitably increasing the potential for research waste. Wasteful research efforts have been reprimanded by scientists for a long time, albeit more strongly in the medical sciences (e.g., Altman, 1994; Chalmers and Glasziou, 2009). Rosengaard et al. (2024) defined five aspects of research waste in biomedical research: methodological, invisible, negligible, underreported, and structural research waste. Methodological research waste concerns study design flaws, poor research conduct and statistical analysis. Invisible waste refers to the premature discontinuation of commenced research, and lack of open science practices. Similarly, underreported waste included poor reporting practices and lack of transparency. Structural and negligible research waste concerns poor planning, low referral to existing research when conducting new trials and unnecessary repetition of research beyond necessary replication efforts. Although this work was conducted on biomedical literature, these categories generally translate to other disciplines (Bowen et al., 2025; Grainger et al., 2020; Miske et al., 2026; Ropovik et al., 2021; Sims et al., 2023).

Research waste is a serious threat to evidence-based practices, as it hinders meaningful knowledge production and translation to practice. To combat this, metascientific practices have been proposed as potential solutions. Although it is challenging to reduce the existence of research waste, evidence synthesis has been one of the suggested methods for reducing the impact of research waste and further proliferation of poor-quality studies. For example, Khalil et al. (2022) have described how scoping reviews can help reduce the research waste by categorizing existing research to identify research gaps, redundancy in research, map existing methodology and aid standardization of terminology within a field. Furthermore, systematic reviews can maximise the value of multiple studies to inform practice and the future direction of primary research. Overall, metascientific research is heavily geared towards reducing research waste and improving scientific methods, which enables identifying evidence-based practices. It is particularly important to adhere to good practices when conducting evidence synthesis, not to further exacerbate research waste by conducting poor-quality syntheses. Therefore, several methods to reduce research waste, such as preregistration of research, data sharing, and adherence to conducting guidelines for evidence synthesis, are present in this dissertation. Open science practices and adherence to conduct guidelines ensure that the evidence synthesis contributes to reducing the impact of research waste.

3.6 Evidence Synthesis

As mentioned in the Evidence-Based Practices in Special Education section, evidence synthesis is the ultimate procedure for identifying evidence-based practices. Although original empirical research is necessary (naturally, there would be nothing to synthesize without it), science is essentially cumulative. No definitive answers can be given only through individual primary studies. Consequently, researchers started collecting multiple findings of the same phenomenon to present generalized conclusions. This procedure has been named in various ways (as things in research usually are), but here I will talk about evidence synthesis. I allow research synthesis as a synonym, but believe systematic reviews and meta-analysis are misnomers, although related concepts. While Chalmers et al. (2002) provide a nice historical overview of the term "research synthesis", for the purposes of establishing EBPs, I will just define it as a systematic identification, appraisal, and integration of evidence from multiple studies to answer a defined research question.

The most important element that evidence synthesis provides compared to other types of evidence collection is the minimization of bias. To establish evidence-based practices, it is imperative that the result stems from a clear path from the intervention of interest to the outcome of interest, without (unknown) interferences to our interpretation of the result, i.e., biases. Evidence pyramids present a hierarchy of evidence, based on level of bias introduced by the study

design. The modified evidence pyramid introduced by Murad et al. (2016) represents study designs in a pyramid form, where the bottom of the pyramid represents high bias designs (e.g., case reports), and the top of the pyramid is reserved for the designs which allow the least biased causal interpretations of results (e.g., randomized controlled trials). The entire pyramid is viewed through the magnifying glass of evidence synthesis, which amplifies the findings of each study design, and is correspondingly of higher evidential value when it includes study designs on top of the pyramid.

There are three levels on which bias can appear and affect evidence synthesis results (see Batinovic et al., 2025, for more information). The first one, within-study bias, is the one mostly covered by the evidence pyramid hierarchy. It most commonly involves bias caused by study design which does not randomize interventions across the sample, low sample sizes or flexibility in analysis choices (see Simmons et al., 2011). Evidence synthesis process reduces within-bias by including higher quality designs and performing risk of bias or quality assessment of studies. This implies that, although not mandatory, high quality evidence syntheses will generally only focus on randomized controlled trials (and potentially quasi-experimental, non-randomized interventions). Secondly, even if only higher quality designs get included, performing a risk of bias assessment and stating the confidence in results consistently with the level of assessed risk of bias ensures negative consequences from within-study bias are attenuated. The final assessment of confidence in the synthesis results is conducted following the GRADE framework (Guyatt et al., 2008), which considers the estimated risk of bias, effect size, inconsistency of effects, appropriateness of outcome measures, and publication bias. The second level concerns publication bias. This bias mostly arises due to publishing traditions, which reward significance and novelty, which skews the publications found in journals towards just barely significant, small-effect-sized, biased literature. Literature bias can mislead the conclusions of the synthesis. To counteract this phenomenon, the synthesis procedure involves searching for "grey" literature and conducting publication bias sensitivity analyses. Grey literature involves reports that are published outside of traditional peer-review publication venues (e.g., preprints, theses, unpublished reports) that are under less pressure to present significant findings. Reviewer bias is the third level, introduced by those who conduct evidence synthesis. These biases can occur in multiple sections of the synthesis procedure. For example, not being transparent about inclusion criteria, being flexible in synthesis models or having only one person screen the literature. Multiple guidelines and handbooks instruct synthesis authors on how to address all three levels of bias (e.g., MECIR, MECCIR, WWC). Additionally, open science practices are crucial for the reduction of the reviewer (synthesis) bias. Preregistering synthesis protocols is mandatory by Cochrane and Campbell. Sharing data templates and search strings to reproduce the search process is equally important. Since data for synthesis is already available and often seen beforehand, adhering to open science practices is evermore important when conducting a research synthesis, to guarantee trustworthy evidence-based practices.

Unfortunately, practices of reducing bias in research synthesis are poorly done in reality. Evidence syntheses in education are regularly conducted without a protocol, without sharing searches or providing extracted data in a FAIR manner, and not conducting any type of risk of bias assessment (Nordström et al., 2023). Furthermore, less than half of meta-analyses in special education included grey literature and only a third examined publication bias according to Gage et al. (2017). These findings are particularly concerning given that head evidence synthesis organizations emphasise the necessity of protocols and sharing practices to ascertain confidence in the evidence. Quality appraisal is the fundamental step in systematic reviews. Ascertaining the quality of synthesized evidence moderates our confidence in the conclusions of the synthesis. Although quality appraisal is not universally defined, the major areas of concern usually covered are quality of the study design, analysis, conduct and reporting of the study. Within-study bias level was mentioned as the first level of bias that can affect evidence synthesis findings. Conducting an assessment of the design, analysis, or reporting standards allows us to estimate the risk of results being biased by potential weaknesses. For example, one of the included studies could be a between-groups intervention design with a large sample size. This

study finds a large positive effect of a reading intervention on decoding ability. A study that has a large sample size will heavily influence the synthesis and skew the final result towards a positive effect. Today, most of these weaknesses can be systematically recognized and appraised with risk-of-bias (e.g., Cochrane Risk of Bias 2 Tool, Sterne et al., 2019) or quality assessment tools (e.g., CEC group design quality indicators, B. G. Cook et al., 2015).

Forensic metascience. The recent explosion of questionable editorial practices, the evolution of paper mills (businesses that mass produce fake/subpar studies and sell authorship; Sanderson, 2024) and predatory journals (journals that employ unethical business practices for-profit purposes), has introduced additional indicators of poor research quality, and even fake studies. Furthermore, plagiarism, low adherence to ethical standards, statistical errors and other reporting inconsistencies are starting to persistently plague the literature. We were parallely introduced to problems that extend beyond questionable research practices (e.g., Stapel incident, Stapel Investigation Committees, Tilburg University, 2012; and the recent Gino fiasco, Datacolada, Simonsohn et al., 2023), and the proliferation of technology (especially with the AI boom) which exacerbated the amount of fake paper and paper mills (Abalkina & Bishop, 2023; Liverpool, 2023) beyond the capacity of individual researchers. Fortunately, there have already been improvements in standardizing the assessment of certain research integrity aspects in medical sciences (INSPECT-SR, J. Wilkinson et al., 2026), which allows development to follow in other fields, and this work is tirelessly continuing by many researchers.

The sub-field of metascience that incepted due to the increased uncovering of research integrity breaches is *forensic metascience*. Forensic metascience "involves using digital tools, statistical observations or human faculties to assess the consistency of empirical features within scientific statements (Heathers, 2025). This area of research has, just like metascience, mostly evolved in the medical sciences, with one of the most prominent advancements being the aforementioned INSPECT-SR tool. Forensic metascience is increasingly becoming a staple in evidence synthesis, as it is no longer enough to evaluate design quality and risk of bias to establish trust in evidence. Not all fields have been catching up with the new trends, regardless of existing needs and challenges. While the medical and psychological sciences have been going through an evolution of crises that has currently culminated into a credibility crisis (i.e., do we even know what we're so poorly researching?; Scheel, 2022), research in education acts aloof toward these threats. Whether that attitude is justified is a different issue. As mentioned, there have been scarce attempts to quantify the issues of replication issues in certain fields within (special) education research, finding the same patterns as other social sciences (cf. Buckley et al., 2023; Perry et al., 2022). Accordingly, open science guidelines have emerged to facilitate improvement in reporting transparency (B. G. Cook et al., 2022; van Dijk et al., 2021). Researchers have further emphasized the implications of open science practices for evidence synthesis (cf. Patall, 2021). However, more serious attempts to address research integrity in education research is still absent, which is a serious shortcoming for evidence synthesis in this field.

Researchers have for decades strove to secure access to adequate and credible education practices for students with intellectual disabilities. Indeed, the importance of access to high-quality education to improve life outcomes cannot be overstated. The advent of evidence-based practices led to the development of rigorous research quality standards that should be adhered to. The process of forming evidence-based practices is currently hindered by several factors. Firstly, the population of students with intellectual disability is complex and diverse, which leads to clear divisions within the special education research, affected by conceptual discrepancies and ontological points of view. Secondly, the landscape of empirical research fails to address the foundational aspects of research – verifiability. These circumstances ultimately contribute to diminished research quality and increased research waste. Several contributors such as inconsistency in transparent reporting or absence of open science practices hinder addressing more abstract inconsistencies that reduce our ability to generalize findings and consider individual needs of students through trustworthy evidence-based practices. This dissertation attempts to contribute to the improvement of state-of-art practices in special education, by utilizing meta-scientific approaches to implement open science practices, improve research assessment and evidence synthesis. The historical developments, disability paradigms and the meta-scientific approach to research for evidence-based practices that have been described in this chapter set the foundation for presenting the open evidence workflow in Chapter 5.

4 Summary of the papers

The aim of this thesis was to gain an understanding of the factors that affect the quality of evidence synthesis in special education research and develop methods to effectively synthesise available evidence, thus improving the quality of evidence synthesis. The thesis is based on four papers, of which three were largely built on the findings from paper I. All data collected and presented from papers I, II and IV is secondary data, and paper III presents preliminary findings from a modified Delphi study and a user pilot. The entire work that formed this thesis can be viewed as a documented and detailed workflow of conducting a living review. I explain the workflow in Chapter 5, and here, I briefly summarize the four included papers.

4.1 Paper I: School-based interventions for primary and secondary school students with intellectual disability: A scoping review

This pre-registered scoping review mapped school-based interventions for k-12 students with intellectual disability. The review had four objectives: to identify types of educational interventions designed for k-12 (primary and secondary school) students with intellectual developmental disorder aim to improve academic, cognitive, behavioural and social outcomes, to identify what research designs were used to investigate the effectiveness of interventions, which measurements were used to assess the effectiveness, and which theoretical frameworks were the basis of identified interventions. A comprehensive systematic search across six databases retrieved 43,483 reports, which resulted in 952 included studies, published 2000-2023. The large set of studies was thematically synthesized into intervention categories. Findings that answered our main objectives revealed that the most common category of interventions were assistive or instructional, and primarily targeted various academic outcomes. Assistive interventions included technological aids, such as tablets or communication devices. Instructional interventions were based on interventions with systematic and explicit teaching methods, such

as direct instruction. Single-case designs were the most prevalent study method, dominated by multiple-probe designs. Furthermore, the studies predominantly measured the outcome with non-standardized measures, which is especially prevalent in single-case design research. We found that the reporting and development of interventions based on theoretical frameworks was rare. Additionally, we shifted from our preregistered plan to conduct study quality assessment of included studies, and instead analysed adherence to open science practices, including data sharing and ethics statement reporting to evaluate research integrity. We found low adherence to open science practices and reporting of ethical statements. The findings highlight the need for improved methods to synthesize prevalent single-case evidence in this field. Overall, the study uncovered a large set of studies with diverse interventions and outcomes of interest. These interventions were largely implemented through single-case designs, and the interventions were primarily based upon previous research, with unclear referral to educational theoretical frameworks. Furthermore, the limited engagement with open science practices and other research integrity indicators warrant further research into methods to evaluate the trustworthiness of identified evidence, and methods for better utilization of available evidence.

4.2 Paper II: Effectiveness of Reading Interventions on Word Recognition for Students with Intellectual Disability: A Systematic Review and IPD Meta-analysis of Single Case Experimental Designs

Paper II is an in-principle accepted Stage 1 registered report. Following the findings from Paper I, we designed a study to systematically review and meta-analyze evidence we found to be prevalent and underutilized. We focused on reading interventions, as reading is an essential skill for participation in society, yet persons with intellectual disabilities (ID) struggle with acquiring functional reading abilities. This systematic review and meta-analysis focused on single-case experimental design (SCED) interventions aiming to enhance word recognition skills among children with ID, reflecting an emergent and important area of research in special education, given the challenges in recruiting large samples. The objectives of this study were to investigate the effectiveness of reading interventions with the primary aim of improving word recognition ability in children with ID, and analyse moderating effects of IQ on the relationship between the reading interventions and word recognition.

The review included intervention studies aimed to improve word recognition abilities among primary school (k-9) students with ID. The eligible interventions are categorized as phonemic awareness, phonics, fluency, and other, and are included only if aimed explicitly at improving word recognition skills. We searched Scopus, Web of Science, APA PsycInfo, APA PsycArticles and Academic Search Complete, and Google Scholar, accompanied with the manual searches of journals. Studies were screened independently by two reviewers. We extracted data on demographics, study design characteristics, intervention details, and outcomes manually in spreadsheet form, while the outcome data for word recognition was extracted from graphs using the juicR data extraction package. Data was synthesized with a Bayesian individual participant data (IPD) meta-analysis. This approach accommodated the expected small sample sizes in SCED research and allowed modeling nuanced criteria of this design to evaluate intervention effectiveness and the exploration of heterogeneity across studies.

Although the Stage 2 procedure is still ongoing, screening of the initial 952 studies from the scoping review is completed and resulted in 43 relevant studies, of which four were meta-analysed in the phonics category of interventions. Updated searches retrieved 16,854 references. The screening phase is still ongoing, and currently resulted in seven included studies, of which one was meta-analysed. Preliminary findings from a meta-analysis of phonics interventions (k=5) show there is a positive, but uncertain effectiveness of reading intervention on word recognition. Furthermore, we were able to learn about heterogeneity on study and participant levels, and

better understand where variation originates. In our preliminary findings, the study-level was larger than participant-level (albeit both levels accompanied with large uncertainty), indicating important contextual factors that could moderate effectiveness of reading interventions. Better reporting standards of methods and materials would facilitate identification of these contextual agents. Further findings suggest poor reporting standards which obfuscate the opportunity to utilize openly shared raw graphical data, and moderate to low study quality of the evaluated studies. These findings further highlight the need to apply open principles to establish credible evidence-based practices.

4.3 Paper III: Developing and Piloting a QUick Evaluation of Research Integrity for Evidence Synthesis in Education (QUERIES-Edu)

The threat to evidence synthesis is becoming more prominent with the increasingly unexpected amounts of retrieved studies, questionable data and inconsistent findings beginning to pollute the publication system and potentially bias meta-analytic conclusions. These concerns were evident in the findings from Paper I. Motivated by the results from the scoping review, Paper III presents preliminary findings from the development of a research integrity assessment checklist: *QUick Evaluation of Research Integrity for Evidence Synthesis in Educational Research; QUERIES-Edu*. This in-principle accepted registered report Stage 1 and an ongoing Stage 2 study consists of a modified Delphi panel of experts who developed a short checklist that evaluates research integrity and trustworthiness. The aim of the study was to develop a checklist that facilitates screening of untrustworthy studies in the early stages of evidence synthesis, thus reducing the time-consuming data extraction and quality assessment, and improving the credibility of the evidence synthesis. The expert panel formed the checklist and accompanying scoring system in three rounds of deliberations and item assessments. The checklist consists of eleven items, and is currently being piloted by researchers from diverse areas of expertise and various career stages. To the best of our knowledge, the checklist presented in this paper is the first attempt to systematize research integrity assessment in educational research, primarily for use in evidence synthesis. Preliminary findings presented in the thesis suggest there is moderate reliability in the assessment of studies. Reliability is similar across three evaluated categories of study designs: randomized-controlled trials, quasi-experimental designs, and single-case experimental designs. However, at this stage in the development process, the perceived usability and accurate detection of problematic studies are at the moment moderate. Further data collection will lead to stronger conclusions about the usability presented in the preliminary results, and illuminate any areas that might need to be revised. The average completion time is promising, being well below thirty minutes across all three categories of designs. Future efforts should focus on developing automated tools to expedite the evaluation, and better forensic methods are needed to address trustworthiness of study designs beyond randomized controlled-trials.

4.4 Paper IV: Living Review Platform for Educational Interventions: The Evidence in Learning and Didactics Community Augmented Meta-analysis

Community-Augmented Meta-Analysis (CAMA) platforms have pioneered the dynamic implementation of living reviews. More directly, CAMAs are a framework for curating evidence synthesis datasets and hosting them online to allow users to interact with the studies, conduct their own meta-analysis, and contribute to the evidence base. Following these principles, we introduce the Evidence in Learning and Didactics Community Augmented Meta Analysis (ELD CAMA). ELD CAMA combines a community based framework with open infrastructure designed to support the long-term development, maintenance, and improvement of open, repro-

ducible, and living evidence synthesis in learning and didactics. We build on existing platforms to address the most pressing gaps, thus creating a database that supports uploading and selecting individual studies instead of complete meta-analyses, and an emphasis on including study quality assessment scores in the meta-analysis. The platform can be accessed here: https://wanda.lnu.se/home. This study describes the newly built platform that supports living meta-analyses of educational interventions for k-12 students and facilitates FAIR standards in evidence synthesis. We provide a breakdown of all components in the cloud-based platform, and a use-case example for uploading, analyzing, and reusing the data from the platform. Finally, we provide implications and future directions for the platform, including ways to adapt it to other areas than evidence in learning. The ELD CAMA represents the final component in establishing and maintaining a base of evidence-based educational practices, which are curated following open science principles, and accessible to researchers and practitioners.

5 Open Evidence Synthesis Workflow

The aim of this thesis was to gain an understanding of the factors that affect the quality of evidence synthesis in special education research and develop methods to effectively synthesize available evidence, thus improving the quality of evidence synthesis. The work following that aim resulted in the development of a workflow for facilitating open evidence synthesis in special education research. The conducted studies represent different stages of the proposed framework. For example, Paper I describes the planning stage of identifying research areas that are well established for a systematic review. This chapter consists of descriptions of findings that inform the development of the workflow, accompanied by my reflections on the experience and the challenges that arose. Additionally, I provide suggestions to help overcome barriers to the challenges in conducting evidence synthesis. These suggestions are highlighted in a "Suggestion box" and are meant to encourage researchers to step into open science, and improve trustworthiness of evidence synthesis in special education research.

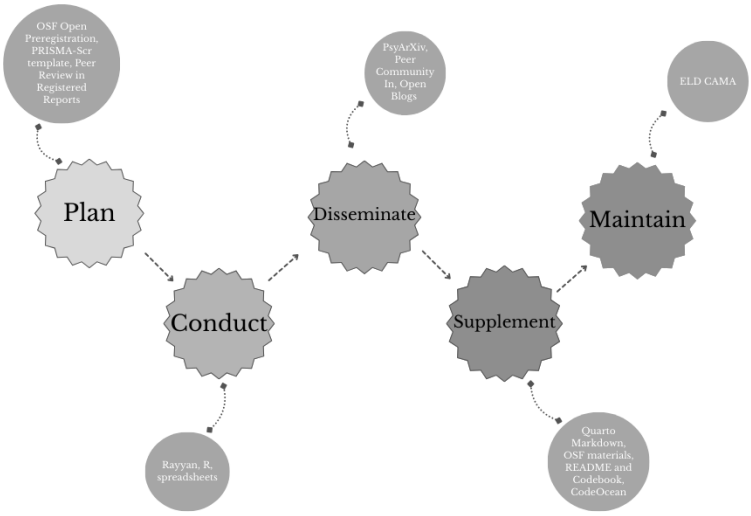
Noteworthy, however, is that the presented work focused solely on quantitative methods, specifically intervention research. Evidence synthesis is possible on non-intervention quantitative methods, and qualitative research. Not all of what I describe is applicable for different study designs, although general open science practices should be relevant regardless of which study design is being synthesized.

5.1 Plan

Protocols

All three empirical studies I present in this dissertation have been pre-registered. Either peer-reviewed (Paper II and III) or not (Paper I), each study had a publicly available protocol specifying the research plan before the studies were conducted.

Figure 5.1: Open Research Workflow for Evidence Synthesis



Conducting scoping reviews begins with developing a protocol where all steps are a-priori decided and made publicly available before conducting the review. Each scoping review should involve specifying inclusion criteria, based on one of the established frameworks. Most commonly, scoping revies follow the PCC (Population-Context-Concept, Peters et al., 2020) framework that requires specifying demographic characteristics (e.g., age, diagnosis, occupation) of the participants, context (e.g., school, low-income countries), and concept (e.g., participation, well-being) of interest. Other frameworks are possible as well, depending on research aims (e.g., PICO(S), Amir-Behghadami and Janati, 2020; SPIDER, Cooke et al., 2012). Based on the inclusion criteria, the protocol should specify sources of literature that will be searched (e.g., citation databases, journals, grey literature) and provide exact search strategy for each selected database. Search strategies should be piloted and relevant. Protocols should also provide information about which items will be extracted from the included studies, describe the screening and data extraction process. Finally, protocols should describe how the results will be presented, which usually involves descriptive statistics (e.g., frequencies), as scoping reviews are not meant for inferential statistics.

For Paper II, we planned to conduct systematic searches of the already included studies in the scoping review from Paper I, together with an updated search from that review. In addition to database searches, there would be a manual search of relevant journals, Journal of Applied Behavior Analysis, Behavior Modification, Research in Developmental Disabilities, and a backward and forward citation search of references from included studies. We would include single-case experimental studies that implemented reading interventions for students with intellectual disability to improve word recognition.

Studies identified through the searches were planned to be screened in Rayyan, and data extracted in spreadsheets and R, using the juicR package (Lajeunesse, 2021). Data analysis is conducted in R, using the brms package (Bürkner, 2017) that provides a Stan (Carpenter et al., 2017) interface in R for Bayesian statistics. Results and analysis are written in a Quarto Markdown file and available on OSF and CodeOcean to promote reproducibility.

Paper I was a simple preregistration, and the other two were registered reports (Paper II and III). Both options had ups and downs. I would say that the higher degree of flexibility and time saving were the only benefits of preregistering my study independently online instead of submitting a registered report. Had I had the time to do things differently, and knew what I know now, I would have conducted all three papers in the form of a registered report. I do not think this should be mandatory for scoping reviews in general, and I believe they fall in the gray line of empirical work (there are no inferences to be drawn from scoping reviews, therefore no need to evaluate the severity of the tests) where exploration is not only allowed, but to an extent desired. However, at the start of my PhD journey, without much insight in the field, I believe peer review of my protocol would've tightened my aims, cleared up the flow, and given me a better overview of the feasibility of this work.

During the planning stage, for the Paper I (scoping review), I should have piloted more. Again, because I was in a cycle of "I don't know much about the topic so I will conduct a scoping review, but I need to know about the topic to plan the scoping review", things caught up with me during the process of conducting the review. Many aspects were pointed out to me by content experts who co-authored the paper, and yet, so many aspects surprised me after all. In hindsight, I feel like I have piloted the data properly only about halfway in the data extraction process (i.e., after doing over 200 extractions), and could at this point do a proper plan to extract that data again. My point is, plan the data extraction section the longest. If you are planning to extract free-text, make sure you standardize every step around it. Keep in mind that you will inevitably have to process the extracted data further and clean it up additionally, so do not tie yourself beforehand into categories. For example, providing a standardized list of possible answers might seem like a good workflow at the beginning to facilitate analysis. In reality, you will encounter multiple answers that fit none of the categories, which you will have to recode later in the process. Rather, develop a plan for recoding the values instead. Search for guides, recommendations for categories, existing categories in published papers. Create scripts and instructions for recoding, write codebook plans. Otherwise, if you are motivated by research questions, you can pre-define categories that will answer your question and limit yourself only to this, whichever does not fit, is not your concern. However, the latter is more applicable to confirmatory research, rather than exploratory, which scoping reviews are. Another benefit of writing protocols is reusability of word formulations, and even content. I was able to reuse and adapt searches, and eligibility criteria from my Paper I in two other protocols. One thing I believe is left to improve for the future is using machine readable templates to pre-register work (e.g., OSF Generalized Review Form, van den Akker et al., 2023) instead of free text documents. I especially recommend against uploading files as PDFs, which I have done. FAIR standards should extend beyond data, and that also applies to preregistration forms.

The other two papers were submitted as registered reports. Timewise, this does not pay off to PhD students, but it is way more worthwhile. Both papers received incredible peer reviewer support. For the meta-analysis paper (Paper II), I've received crucial comments that improved accuracy and flow of the analysis. For a synthesis that has not yet been used and done like this, comments from experts before doing the study were essential. They guided me towards more feasible options, and allowed me to fix things before it was too late. Although doing registered reports requires more effort during the planning stage, it is also helpful when you get to a point where you are stuck. The best course of action is to do the plan as detailed as you can, then submit for review, where peer-reviewers can alleviate you from making the impossible choices.

All in all, planning your research to the point of making it worthy of publishing before it is conducted is difficult. It is also necessary. But rewarding too. Thinking ahead is the only safe

way to walk through the garden of forking paths. It is the perfect time to go over the what-ifs. It helps develop researcher skills, think about causal effects, and possible confounders. It forces everyone to specify a clear research question and a reasonable way to answer it. All of the components of research which are not done consistently well (Scheel, 2022), but are the core of science.

Paper I: Scope it out

Paper I was a scoping review of school-based interventions for students with intellectual disability. Scoping reviews are types of systematic literature reviews, first introduced by Arksey and O'Malley (2005), and developed over the years into a consolidated framework today best explained by JBI (Aromataris et al., 2024). These reviews are used to answer broader questions about the field of interest, explore the breadth of the conducted research and clarify concepts. Unlike systematic reviews, scoping reviews cannot and should not answer precise and narrow questions about causal evidence or magnitude of relationships between phenomena of interest. Scoping reviews are precursors to systematic reviews, and can be utilized to develop a database of studies eligible for a systematic review. Unless there is a commissioned need for a specific systematic review, scoping reviews should be the natural first step towards systematic reviews, to decrease research waste. There are similar review types that focus on exploratory research questions and diverse studies, like evidence gap maps and umbrella reviews. These reviews usually also search for other reviews, while scoping reviews mostly focus on original studies. Each review type serves a different purpose, and I selected a scoping review for the first study as it best fits my objectives – curate a collection of primary school-intervention studies and identify patterns in employed study designs, measured outcomes and theoretical frameworks that direct the study.

For my scoping review, I followed JBI guidelines for conducting scoping reviews, both for the development of the protocol and the data collection. I also followed PRISMA-ScR reporting guidelines for reporting the conducted review, and PRISMA-ScR and PRISMA-P reporting guidelines when writing the protocol.

Eligibility criteria for this review was intentionally broad, and included any school-based intervention for primary and secondary students with intellectual disability (ID). Outcomes could have been targeting any area, and we included any experimental design, as long as there was an intervention implemented on students with ID. To be included, studies needed to report IQ of the student (lower than 75) and/or state that the students had intellectual disability. Eligibility criteria followed the PICOS[1] (Population-Intervention-Comparator-Outcome-Study design) framework as the aim of the review was to curate a dataset of studies for further systematic reviews.

We conducted a comprehensive search in six databases (APA PsycInfo, APA PsycArticles, Academic Search Complete, Web of Science, SCOPUS, ERIC) and Google Scholar, with an exhaustive search strategy to uncover all existing school interventions. The searches covered peer-reviewed and grey literature, and were conducted in August, 2022, and updated in November, 2023. Additionally, we screened references of systematic reviews that were identified through our database search for potentially missed eligible studies. Each study was screened by two reviewers with a software for conducting reviews, Covidence. Data extraction was conducted by one extractor, with 25% extractions validated by a second extractor. Extracted studies were descriptively analyzed in R, and data analysis and the result section were written in Quarto Markdown to streamline reproducibility and updating.

[1] PICOS — Population, Intervention, Comparison, Study type — is a framework for describing eligibility criteria for studies to be included in a systematic review. Created to provide guidance for intervention research, it allows reviewers to transparently describe the most important components of a study which can affect synthesis results.

Searches yielded over twenty thousand articles to screen, after deduplication. A true nightmare. We expected many studies because we deliberately made the searches highly sensitive and not specific. We had some insight into the field and its scattered nature, but we were still caught by surprise. The abstract screening phase was not too daunting, we were more lenient in accepting anything that was not clearly eligible. Then came full-text screening which met us with over 3,000 articles to read. This process was rather exhausting. It was also quite beneficial. I was forced to learn about the field in the most rigorous and systematic way. I'm not sure there is a better way to get a PhD candidate to read hundreds of articles from their field to build a strong knowledge foundation. After this stage, I had a much better mental picture of the research traditions in the field of special education.

Data extraction was done on 952 studies from 940 reports. This stage was extremely challenging, beyond our planned process for extraction. Not only were there hundreds of articles, information we planned to extract was often missing, fragmented, or unclear. At this point, we realized our piloting exercise during the planning did not suffice for the influx of diverse studies we had to extract. Initial extractions were difficult, and long, taking on average an hour and a half for one article. This was mostly due to the discrepancy between what was truly reported, and my insecurity and belief I just was not capable of identifying information. With experience, some 100 reports in the process, I began to see patterns in reporting between different publishers, journals, and types of sources. My average time went down to 30 minutes for simplest articles, and 45 minutes for longer reports. Dissertations and journal articles both had advantages and disadvantages. Dissertations were longer and overwhelming, but often provided all necessary information in greater detail. Conversely, journal articles were digestible, but sometimes lacked information due to their, often journal-imposed, brevity.

Overall, demographic information was more straightforward to extract, while intervention and outcome characteristics took the longest. Generally, there rarely was reporting that leads readers from the research questions to the results explicitly, i.e., aim → research question → intervention and outcome → operationalization of the independent and dependent variable → analysis plan. This led to data extraction being more challenging than was planned in the protocol, because I expected studies to follow this structure when reporting the research, which does not appear to be the traditional reporting system in this area of disability research. These findings generalize Scheel's (2022) assertions about underdefining the scientific claims beyond psychological research. However, intervention procedures and accompanying materials were often described in more detail, albeit the quality of the descriptions could only be evaluated by attempting to replicate the findings, which was out of scope for this thesis.

Suggestion Box

- Analysis: Including code into your preregistration allows you to be realistic about your analysis → create mock data and try to code the analysis plan to understand the feasibility of your design
- Timeline: expect that your research will take longer overall, but the better you plan, the easier it will be to conduct the study
- Experience: with more preregistrations, planning skills become better → plans become more realistic and planning becomes more efficient

5.2 Conduct

Paper II: Systematically search and meta-analyze

Paper II is built on the findings from Paper I, and is a systematic review and meta-analysis of reading interventions. The Cochrane Collaboration, the highest authority for medical evidence

synthesis, states that systematic reviews "seek to collate evidence that fits pre-specified eligibility criteria in order to answer a specific research question. They aim to minimize bias by using explicit, systematic methods documented in advance with a protocol" (Cumpston et al., 2023). Therefore, to answer a precise question about causal effects of reading interventions, studied with single-case experimental design, we planned a systematic review and a meta-analysis. We followed conducting guidelines from the Cochrane Collaboration sibling organization, Campbell Collaboration, which provides guidance for non-medical evidence syntheses. Systematic reviews are not completely dissimilar to scoping reviews. Their scope is usually narrower, and research questions usually require inferential analysis. Additional to the steps required for a scoping review, a crucial component of systematic reviews is a risk of bias (or study quality) assessment. Furthermore, systematic reviews are generally conducted with the goal of meta-analyzing the identified evidence to provide conclusions about the effect of interest.

Meta-analysis of studies located through a proper systematic review provides the least biased, strongest indication of the intervention effect magnitude. Essentially, meta-analysis creates a weighted mean average of effect sizes from each included study. One of the earliest known data synthesis methods was introduced by Ronald Fisher, who developed a technique for combining results of independent hypothesis tests. The technique is considered an analysis of analyses, i.e., meta-analysis. However, the meta-analytic method most commonly referred today as aggregation of results from several studies was introduced by Karl Pearson in 1904, and was formally developed and named by Gene Glass in 1976. Meta-analysis allows us to generate more precise effect size estimates (due to increased sample size from combining studies), resolve conflicting findings from different studies, and generalize answers across populations. At its simplest, meta-analysis is conducted by calculating the same effect size for outcomes of each included study, weighting each effect size by its inverse variance (which gives more weight to more precise estimates) and dividing the combined weighted averages by the combined studies' weights. Throughout years, meta-analysis has been developed into more complex analyses and simultaneously became misused to the point of devaluing its utility and importance as an analytical tool (Kadlec et al., 2023; Lakens et al., 2016).

Generally, development of meta-analysis focused on causal effects investigated through group experimental designs. Today, we have various approaches to this analysis, discernment of clustering issues, diverse multilevel and network models, and accompanying sensitivity and bias analyses. However, the field of single-case experimental design is still catching up to this method. As single-case experimental designs (SCED) are new to quantitative analyses to begin with, it is not unexpected that meta-analytic modeling development is not a priority in this area. Some newer contributions in the area focused on development of standardized effect sizes, particularly those that could be compared to effect sizes from group designs, for example the standardized mean difference (e.g., Chen et al., 2023). Although this approach might yield valid results in certain cases, such as with continuous outcomes, single-case experimental designs are fundamentally different from group designs. SCEDs often focus on a few cases, usually with repetitive outcomes that are in the form of counts or proportions, and the design can be characterised as an interrupted time-series design. Furthermore, the combination of the measure and the design can lead to floor and ceiling effects, which are better described with different distributions. These features make SCEDs so different from conventional group experimental designs, that trying to compare intervention effects seems meaningless.

Firstly, SCEDs have different strengths that should be utilized. SCEDs reports by default provide raw data, albeit in graphical forms, which is a rarity in group design studies. Access to raw data enables us to conduct individual participant data (IPD) meta-analysis. IPD meta-analysis can be done in one or two stages. One stage IPD meta-analysis is essentially identical to a multilevel random-effects regression, clustered by studies (Riley et al., 2023). Although the common meta-analysis can generally be considered a multilevel model clustered by studies, IPD meta-analysis provides higher precision as we have access to raw data points for each participant instead of average effect sizes with reduced variance. Secondly, while we can exploit the benefits of

shared raw data to conduct a more precise IPD meta-analysis, unlike group-designs, single-case experimental designs include only a few participants (cases). Meta-analysis in itself provides better precision due to combining participants from multiple studies, but this might still provide imprecise estimates. However, a Bayesian approach can remedy this disadvantage.

Sample size is important for precise estimates of the parameters. They are even more crucial in complex models and hierarchical data structures. Bayesian statistical inference involves fitting a probability model to a set of data and presenting the result as a probability distribution of the model parameters (see Gelman et al., 2013). Unlike frequentist statistical approaches, the Bayesian approach allows increasing precision of estimates through the introduction of prior knowledge directly in the estimation process (Wong, 2025). Prior knowledge narrows the range of plausible parameters, and increases precision even in small-N cases. Therefore, conducting a Bayesian meta-analysis can help us take advantage of small sample sizes. Finally, unlike meta-analyses of effect sizes from group designs (both within- and between-group experiments), SCED effect sizes have less reliable sampling variance, which is affected by auto-correlation present due to the design type. There have been attempts to develop workarounds, with varying success depending on the effect size (Chen & Pustejovsky, 2024), but more promising approaches point to variants of logistic or linear regression models, depending on outcome type (Grekov et al., 2025).

In summary, Paper II aimed to address a gap in open and reproducible evidence synthesis of primary studies with SCED design. SCEDs have been overwhelmingly prevalent in researching intervention effects for students with intellectual disability (as evident from Paper I), and require thoughtful and reproducible evidence synthesis. Paper II attempts to synthesize the effects of reading interventions on word recognition by conducting a one stage, individual participant data Bayesian beta-binomial regression. We follow Campbell Collaboration Conducting Guidelines (MECCIR; The Methods Group of the Campbell Collaboration, 2016) and the PRISMA 2020 (Page et al., 2021) reporting guidelines, and where applicable, extension for individual participant data meta-analysis reporting guidelines (Stewart et al., 2015). Preliminary findings of the ongoing Stage 2 identified five studies that could be meta-analyzed as one category of interventions with similar employed designs. The general screening procedure was expedited by the already conducted scoping review, where the study characteristics have been previously extracted. Updated searches are currently being screened. Some preliminary results are described to explain the evaluation of evidence discovered during the conduct stage further in this chapter.

Suggestion Box

- Extract studies in batches by category (e.g., from the same author, same journal, or same publisher) → it is faster to extract data reported in the same format, and it is easier to notice patterns in the reports.
- Practicing data extraction (i.e., piloting) improves future data extraction plans → understanding the literature facilitates clearer definitions, rules, and limitations, which in turn makes double-coding more reliable.
- Scoping review data extraction benefits more from consensus between extractors than independence of coding → gaining mutual understanding of the studies is part of the scoping process, as long as you preregister what you are planning to extract.

Paper III: Check the quality

Our first preregistered protocol for the scoping review (Paper I) stated we would conduct a quality appraisal of included studies. During the data extraction process in the scoping review, we started recognizing these issues that pointed to research integrity disregards. As described in the introduction, we recognized the potential threat to evidence synthesis these papers might

pose if left unaddressed. Thus, we deviated from the protocol and decided to systematically analyze the literature for red flags. The issue arose as we did not have a standardized checklist that we could implement on this literature.

Therefore, the third paper was a development study for a checklist to quickly evaluate research integrity in evidence synthesis of educational research (QUERIES-Edu). Our aim was to develop a checklist that is short and simple to implement and can help quickly weed out red flags in papers to remove them from further steps in evidence synthesis – quality assessment and meta-analysis. The items in the checklist would be directed towards identifying potential plagiarism, research integrity mishaps, issues with reporting consistency or implausibility of the reported findings. Ensuring that the study has not breached any major integrity signals allows reviewers to more confidently assess the quality of the reported study, while trusting the reported information is reliable.

A modified Delphi method was used to develop the checklist. The Delphi technique is a well-known, conventional method for collecting opinions and establishing consensus on various decision-making frameworks in educational research (Clayton, 1997). Said opinions are most often gathered by experts, but can also involve a more diverse set of participants. The procedure involves gathering experts or other stakeholders in the field, specializing in relevant practices, and collecting their opinions through surveys or other communication means. Generally, this process requires multiple rounds of iteration to achieve consensus on the matter. In the first phase, the expert panel receives a set of items–statements to rate numerically and provide feedback. Once all panelists provide their comments, project coordinators summarize the ratings and feedback, and revise the items accordingly. The revised list is then sent back to the panelists for further evaluation. This process is then reiterated several times until a consensus is reached (Clayton, 1997).

The Delphi method has become the prevalent method of reaching consensus in healthcare research and shareholder policy projects. The Delphi method capitalizes on the expertise of participants which is especially useful in developing new principles, checklists or guidelines, for which there is not much empirical data. Naturally, the method is not flawless, and critiques around its validity and reliability have been considered and discussed for decades (e.g., Diamond et al., 2014; Goodman, 1987; Khodyakov et al., 2023). The necessary sample size is not established, and the selection process is by design biased – panelists are selected based on their characteristics (i.e., expertise), not randomly. The selection procedure could affect the results in multiple ways: low replicability, susceptibility to the partiality of the panelists, actual expertise of the panelists, and the determined process of reaching consensus. Although Delphi panels do not represent unbiased data collection that research strives for, this method provides a beneficial first step in identifying and defining the scope of knowledge relevant to the topic of interest. Therefore, a modified Delphi method was considered the best approach to developing a research integrity tool for evidence synthesis of educational interventions.

The QUERIES-Edu checklist was developed with two Delphi panels. The first panel consisted of experts in evidence synthesis, forensic meta-science, reproducibility and/or open science practices. These panelists were required to provide numerical ratings on clarity, importance and relevance of presented items for evaluating research integrity. The first two rounds of feedback were collected through written surveys, while the final round involved a roundtable meeting with the project leaders and expert panelists to reach consensus on the items, and accompanying guide. Once the checklist was finalized, the second panel of user panelists received the checklist to pilot on a set of papers selected by the project leaders. These papers were selected from the dataset curated in Paper I, and represented educational interventions implemented with various group and single-case experimental designs. The second panel included individuals with varying degrees of experience in conducting evidence synthesis, and in different career stages, from early career to senior researchers. In this phase, we aimed for a diverse set of researchers to test the checklist on the papers to assess how applicable it is, depending on experience with research synthesis and overall career. The user panelists had to evaluate 22 papers and rate

Table 5.1: QUERIES-Edu checklist items

Item Nr.	Domain	Item
0	Integrity Validity	The journal is classified as having a functional peer-review and is not considered a predatory journal.
1	Integrity Validity	Has the study been retracted?
2	Integrity Validity	Does the study have a particularly alarming expression of concern, a critical Retraction Watch coverage or a concerning PubPeer comment?
3	Consistency	Are there any discrepancies in the data reported in figures, tables or text?
4	Consistency	Is there evidence of copied or manipulated work within paper, such as duplicated or partially duplicated tables or figures/images?
5	Transparency	Is there insufficient information reported to assess the plausibility of the study (e.g., country, participant demographic information, study conductor, procedure description)?
6	Plausibility	Does it seem implausible for the study to be completed as described?
7	Consistency	Are details in the publication inconsistent with those in the registration documents?
8	Consistency	Are there large sections of text that are nonsensical or incoherent?
9	Plagiarism	Did you notice any tortured phrases?
10	Overall Assessment	In your overall assessment, are there too many 'pass with concern' to confidently include the study?

Note. Presented items are preliminary results, and are in the process of being tested by the user panel during the time of writing this dissertation.

the checklist's usability. Project's coordinators would summarize the findings, and if necessary, revise the checklist together with the expert panel and complete the checklist development.

Table 5.1 lists the selected items and the domain of research currently evaluated by the user panel. The checklist attempts to guide the readers towards the most relevant problems in a study that can easily be spotted through a skim of the methods and the result section. With this workflow, studies that do not present trustworthy results can be screened out before additional effort in their assessment is invested by researchers. For example, a study might have duplicated table results, or mismatched results reported in text and tables. Regardless of the source, the checklist is designed to elucidate areas in the paper which reduce trust in the presented results, ultimately allowing researchers to exclude potentially erroneous results from the meta-analysis. Preliminary findings from nine users indicate there is moderate reliability in the assessment of studies across all evaluated designs (randomized-controlled trials), quasi-experimental designs, and even single-case experimental designs. The latter designs have not been under as much scrutiny as randomized-controlled trials (particularly evaluated in medical fields; INSPECT-SR). Nonetheless, a larger sample of users is needed to conclude whether the checklist can reliably evaluate all relevant designs prevalent in special education research.

Evaluate Evidence

The findings from studies presented in this thesis have illuminated areas within research designed to establish evidence-based practice that I consider imperative to describe in more detail. The following section draws upon the findings to guide evaluation of evidence, intended for researchers and practitioners who find special education and evidence synthesis relevant to them.

Scientific literacy is a crucial skill for combating misinformation, especially today, when science is manipulated and taken hostage for malevolent purposes. However, scientific literacy should not be exclusive to scientists, but available to practitioners and laypeople relying on the evidence as well. Guidance for understanding and implementing evidence based practices is left to policy makers and institutions which make generalized decisions on school, region, or country level. Although practitioners (in this case teachers, educators, paraprofessionals or clinical workers) can rarely practice against established recommendations, it is still beneficial to understand evidence straight from the source – scientific reports. This section provides a guide towards understanding the quality of evidence synthesis and the relevance of its findings. Explanations in this section are based on the findings from this thesis, and are mostly geared to disability practitioners working in schools with individuals with intellectual disability. The chapter can also provide insights to evidence synthesis for early career researchers or disability researchers unfamiliar with evidence synthesis. Lastly, this section describes in more detail the important features of the summary planned to be presented on the CAMA platform discussed in the earlier section (i.e., Figure 5.7).

How to trust the evidence?

Paper I revealed issues and patterns in this literature that should be taken into consideration when reading findings from evidence syntheses in this field. The following recommendations are based on a framework developed to assess the confidence in the meta-analytic results–GRADE framework. If a synthesis presents findings following the GRADE system, this interpretation of results is more likely to be trustworthy. However, in the case a synthesis does not follow the GRADE framework, which is more likely in fields outside of medicine, this section will provide recommendations for readers to do their own GRADE assessment and estimate how much they trust the presented evidence. There are five domains of the framework: limitation in study design and execution, dissemination bias, inconsistency, imprecision, and indirectness.

Limitation in study design and execution. Firstly, practitioners reading evidence synthesis should look for mentions of "quality appraisal", "risk of bias assessment", or other headings signaling that researchers evaluated the quality of the designs in the included studies, and appropriately modified their analyses to reflect the evaluations. This means that the result section should mention moderator or subgroup analysis which either exclude poor studies, or a narrative reflection on the findings given the quality assessment. Risk of bias is essential, and I can confidently recommend practitioners to completely disregard systematic reviews and meta-analyses that omit risk of bias assessments. One can of course evaluate the papers independently and decide if they trust the presented evidence based on their evaluations, but practitioners should not be required to conduct additional research to comfortably trust the presented findings.

However, with the technological developments, and the proliferation of research findings, it is no longer enough to just conduct risk of bias assessments. As findings in Paper I show, the amount of educational intervention studies in the field of intellectual disability research is overwhelming for synthesists. Filtering through these studies is resource-intensive, and requires deeper verification procedures than assessment of design quality. Therefore, syntheses that do not also engage with research integrity evaluations of included studies should be seen as potentially untrustworthy. Any newer systematic reviews should provide information about screening for problematic trials, by using more established tools, such as INSPECT-SR (J. Wilkinson et

al., 2026) or TRACT (Mol et al., 2023), or newer, education-appropriate tools like QUERIES-Edu developed in Paper III. For example, QUERIES-Edu items guide readers to evaluate the article for any glaring impossible statistical values, discrepancies between values reported in different areas of the article, lack of information needed to evaluate plausibility, or implausible methods and findings (Table 5.1).

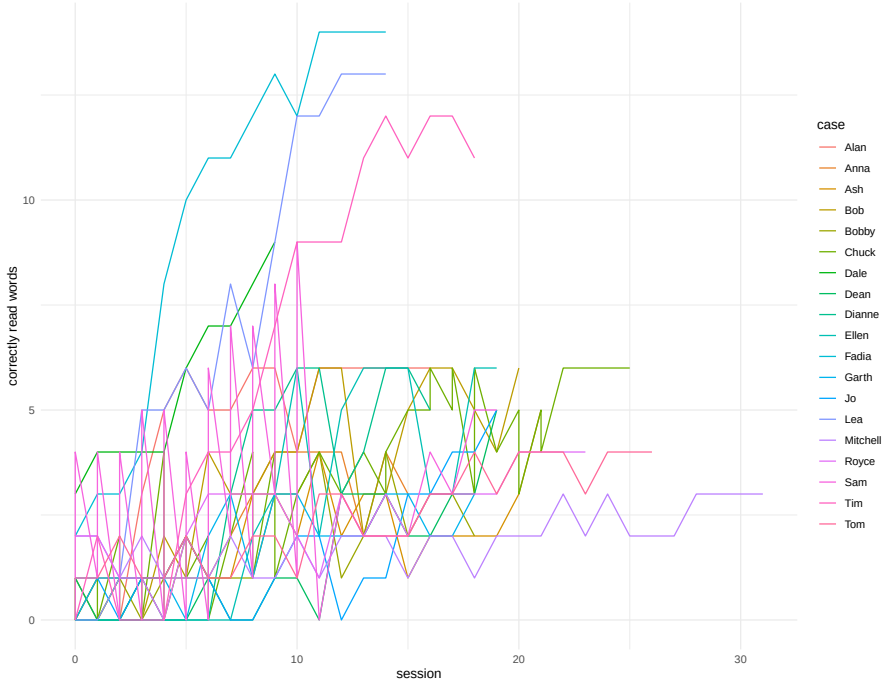
Lastly, research integrity should not only be evaluated by the researchers, it should be present in their synthesis as well. This implies that the review follows open and reproducible research practices. This can easily be checked at the beginning of methods sections, or under Data Availability Statement headings, where the authors will clearly state whether they preregistered their protocol online, and shared data and code used to produce the synthesis. Although the open science movement might not have been prominent in the educational sciences before the recent five years (i.e., before 2021, which is generous in itself), synthesis guidelines have long proposed sharing protocols of reviews. Therefore, unavailable protocols can safely be considered a problem, even for reviews that were published before the open science boom.

Dissemination bias. In the same line, evidence synthesis should assess publication bias (when this is possible, single case research so far has not developed reliable publication bias assessments). As described in the introduction, publication bias reveals how much the field suffers from pressure to only publish significant findings. This indicates an imbalance in available evidence which can falsely drive the meta-analytic results toward significant effects. Although evaluating publication bias is a complex task even for experienced researchers, practitioners can look for mentions of "grey", "non peer-reviewed" or "unpublished" reports included in the evidence synthesis, which signals that the synthesists attempted to minimize potential publication bias by including findings that might be less likely to report desirable findings.

Inconsistency. Synthesists should evaluate inconsistency of evidence. Inconsistency means that the included studies differ, and this inconsistency can be low or high. Naturally, it is expected that most syntheses will find inconsistency in their studies. It is important that they address it. Inconsistency can occur in two ways: by random measurement error, or by true heterogeneity between studies. Random measurement error is not something that can be corrected, and if it is high, indicates that the studies are poorly done, and should not be trusted. On the other hand, true heterogeneity signals systematic differences in the studies which indicate that they should perhaps not be synthesized together. High heterogeneity can be "solved" by subgrouping studies into more comparable categories and synthesizing them again. These categories should preferably be predefined and justified appropriately. If heterogeneity remains and cannot be clearly resolved, this evidence should once again not be trusted as informative. This information is usually found in statistical reports of variance (expressed as τ^2 which is the variance in meta-analyses), and percentage of true heterogeneity (which is called I^2 statistic). Researchers should elaborate on the size of the variance (which presents the overall variation of studies, combining measurement error and true heterogeneity), and the percentage of heterogeneity. Practitioners should not have to understand the values and interpret them on their own. High quality evidence syntheses would properly interpret both the size of variance (e.g., τ^2) and heterogeneity (I^2) in relation to each other, along with the Q statistic which simply identifies whether the heterogeneity is statistically significant. Often, researchers will only interpret one indicator of heterogeneity, most often the I^2 measure as it is presented on the most understandable scale. However, interpreting only one indicator in isolation is wrong and can be misleading, and readers can consider that a signal for lower confidence in the presented findings.

For example, findings could find a low overall variance, and high heterogeneity. These findings would imply that the synthesized studies do in fact systematically differ, but if the effect size is large, due to low variance we can conclude that those variations are practically inconsequential for the findings. However, if researchers only report I^2, i.e., heterogeneity, this would falsely lead readers to believe that these synthesized studies differ too much. Figure 5.2 from Paper II illustrates variability of trajectories between individuals from different studies, and within those

Figure 5.2: Individual trajectories of participants in reading intervention studies

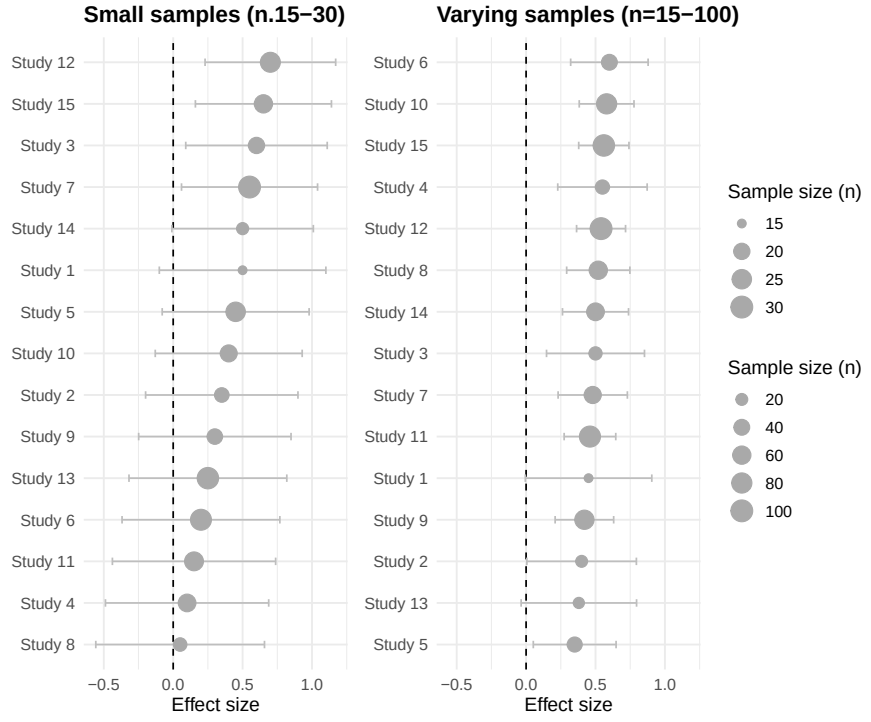


Note. The figure is based on the preliminary data and reproduced from Paper II, under CC-by 4.0 license.

studies. Although all participants received a phonics intervention, were tested on decoding ability of words, and participated in a multiple-baseline or probe design, there are still several variations, such as intervention length or different contextual factors.

Imprecision. Imprecision refers to the exact estimation of the intervention effect when conducting a meta-analysis. As we cannot precisely estimate an effect (as estimating something 100% certainly would be an oxymoron), statisticians have thought of a way to ballpark that value as close as possible. The basic assumption of this estimation includes a scenario where one measurement is conducted an infinitive amount of times, in an infinitive number of new studies, and produces an interval around the mean estimate which spans across 1.96 standard errors in negative and positive direction. These intervals are expected to contain the true population value of the phenomenon of interest (i.e., the estimand) in 95% of the conducted studies. This estimation method is called the 95% confidence interval. However, as nobody will ever do an infinite number of studies, a practical interpretation can be that there is a true value contained in this interval with 95% confidence. This means that wider confidence intervals imply higher imprecision. For example, an effect of scaffolding on learning to multiply could be g= 0.5, which is colloquially considered a medium effect size. However, if a confidence interval spans from g= -0.1 to g= 1.1, the true value could be negative (implying that scaffolding negatively affects learning multiplication). That level of imprecision renders our mean estimates less informative, or sometimes even useless. Figure 5.3 shows two alternative forest plots. The black dashed vertical line represents no effect (e.g., no difference between scaffolding intervention and usual teaching). Gray circles represent the effect size of each study, with the size of the circles cor-

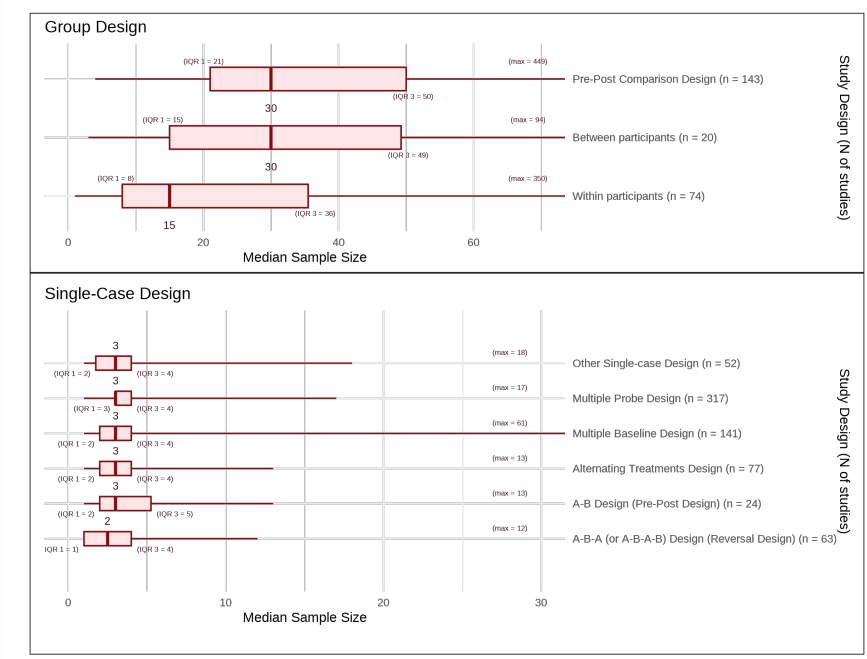
Figure 5.3: Example forest plots with narrow and wide confidence intervals.



responding to the sample size of the study. The horizontal gray lines are confidence intervals around each estimate. The forest plot on the right shows narrower confidence intervals than the plot on the left. The confidence intervals on the left plot cross the vertical dashed lines often, implying that, although the estimates (circles) represent varying strength of the intervention, we cannot be certain that there truly is an effect.

Imprecision is most often affected by the sample size, with other sources of measurement error adding to that imprecision. From the results of the scoping review, both group and single-case designs had small sample sizes, but this is particularly problematic for group designs (Figure 4 in Paper I). Generally, group designs sampled around 15 students, which is often too underpowered to detect true effects, unless they are incredibly large—which is not a common case in educational interventions. Although meta-analysis is specifically meant to increase power by combining multiple studies and thus increasing the sample size, if all included studies are small and report inconsistent effect sizes, this can still produce wide confidence intervals. Additional causes of imprecision are related to the variability of the sample. As results of the scoping review show, sample variability too is a potential issue in this literature. Diagnostic measures are heterogeneous, often missing standardised diagnostic test results. Additionally, there is a variability in intellectual disability severity, and co-occurring conditions which can be relevant for learning abilities (e.g., neurodevelopmental disorders, genetic disorders). Therefore, it is important to expect variability in results of systematic reviews that synthesize evidence from this literature. Even more importantly, interpretations of the results in the review should only be

Figure 5.4: Average sample sizes per study design



Note. Figure reproduced from Paper I, published under CC-by 4.0 license.

trusted if the authors transparently report imprecision and make statements which appropriately integrate imprecision in their assessment of the certainty.

Indirectness. The final indicator of certainty in evidence is the indirectness of collected evidence. Indirectness affects applicability and generalisability, among others, and is evaluated by comparing study characteristics (i.e., PICOS) with the review research question. To keep the example consistent with the previous paragraph, let us say the research question of interest is "How effective is instructional scaffolding for teaching multiplication to 8-year-old students with intellectual disability?" When we identify studies that can answer this question, indirectness could come from the intervention characteristics, variety of outcome measures, sample variability or context of the interventions. Based on the results from the scoping review, interventions that implemented scaffolding had different "doses", i.e., the intervention was implemented a different number of times, broken down in different steps between studies. Additionally, multiplication outcomes could be measured with almost as many different tests as there were studies. Tests could be bespoke, measuring the tasks with multiple-choice or free text alternatives, standardised tests like KeyMath, or a combination of observations. Finally, populations could vary, from mixed samples, students with co-occurring disabilities, or varying severities of intellectual disabilities. These differences in how original studies were conducted prevent reviews from making direct observations of the effect of the intervention on the outcomes of interest. For example, we cannot know if the effect of the interventions affect the ability to recognise the correct multiplication (e.g., through a multiple-choice task) or correctly multiplying by recall (free text answers). We further do not know certainly if the effects are as strong for students with moderate intellectual disability if they do not have co-occurring learning disabilities.

Reviews should not have issues with indirectness generally. They should reduce the risk of it by having highly specific PICOS requirements. For example, if a reviewer is interested in the effects of reading interventions on seven year-old students with ID, they can mitigate indirect evidence by narrowing the eligible population to seven year-old students with diagnosed intellectual disability. Further, they can decide to only include standardised tests they predefined in their protocol, with specific components each intervention should contain. However, if the review still has a broader inclusion criteria, with the aim of assessing potential generalisable effects, they should report moderator analyses to confirm that decision was appropriate. Reviews should justify why they conducted a moderator analysis (which they usually do when there is high heterogeneity, i.e., the I^2 value). Additionally, the moderator (or subgroup) analyses should be properly powered by a high number of studies to be informative. The review should report whether the indirectness of the evidence changes the outcome, if the moderator analysis shows that effects differ based on different characteristics of the outcome, intervention or the population. To conclude, if the authors cannot claim that the included studies represent direct evidence for the effect of the intervention on the outcome, then their certainty in the meta-analytic findings should be considered lowered due to indirectness of evidence.

Overall, GRADE components can simply guide the readers through the findings of reviews, and allow them to evaluate the certainty of evidence presented for a practice of interest. Although high quality reviews will provide this assessment for the readers, when evidence syntheses do not evaluate certainty of evidence, interested practitioners can read the reviews guided by this framework.

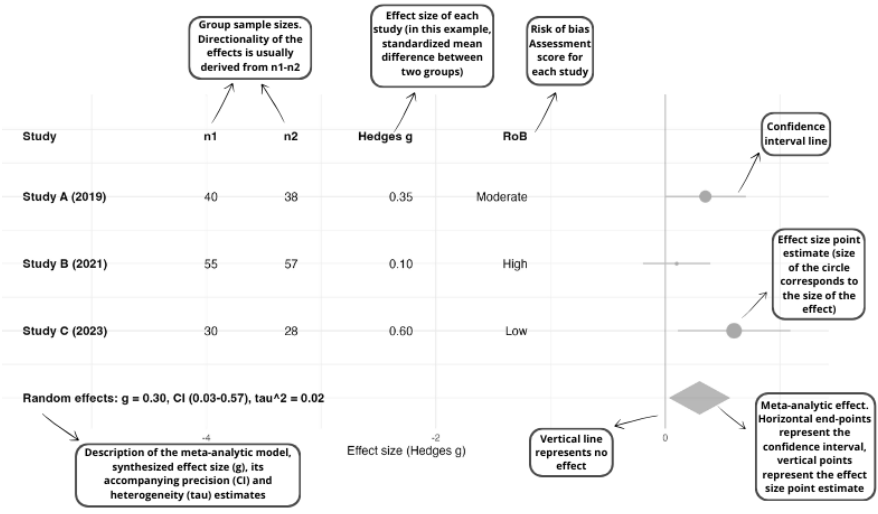
How to understand the evidence?

It might be argued that understanding whether evidence is compelling should come before evaluating if it is trustworthy, but I believe that all evidence is compelling if it is trustworthy, so this section comes after learning how to evaluate where the results are credible. In general, evidence synthesis of intervention research aims to establish certain evidence on whether an intervention of interest works. Although the most common question is what works, a high-quality evidence synthesis also provides insights into what works for whom. This is accomplished by finding all relevant intervention studies, extracting their statistical results, and combining these results into one finding that is more robust and informative than any individual study in the synthesis.

Effect size. The core of the synthesis procedure is the effect size. An effect size is the statistical representation of the magnitude of intervention's effect on the outcome. For example, one study included in Paper I examined the effectiveness of a comprehensive reading intervention for improving receptive vocabulary in students with moderate intellectual disability. A way to test this is to form one group of students who will receive this reading intervention, and one group that will receive usual teaching methods. After the intervention has been delivered, both groups of students will be tested with the same test. We might be interested in whether these two groups differ statistically significantly, but more importantly, we should be interested in how much the two groups differ. The latter is expressed with effect sizes. In the described example, the most simple effect size is the difference between mean scores of the two groups. The intervention group could on average score an 75 on the Peabody Picture Vocabulary Test (a 100 represents an average score for the typically developing students on a standardized scale), while the control group might on average score 67. Depending on how many participants are in the group, this difference can be statistically significant or not. The unadjusted mean difference will however always be 8 points. On average, the intervention group scores 8 points more on the test. However, comparing mean differences is too simplistic and can skew the real differences. It is possible that the two groups have different ranges of values. Thus, it is common to standardize the mean difference to interpret the effect size as the difference in standard deviations. This means that researchers will most commonly report standardized mean differences (SMD; often in the form of Cohen's d or Hedges' g values). SMD is calculated by subtracting the means and dividing that difference by their pooled standard deviation.

To illustrate this, the Peabody Picture Vocabulary Test [PPVT] is a good example. I will use the scoring as a pedagogical example, although this will in practice differ, as we calculate the SMD statistically differently than how I explain it here. As mentioned, PPVT already has a standardized scoring system, where an average reader scores 100 points. In addition to this, the average range is 100 +- 15 scores, and 15 represents the standard deviation of scores on the test. If the example group mean difference is 8 points, this is slightly over a half of the standard deviation of 15, or 0.53 standard deviations. A common rule of thumb interpretation introduced by Cohen (Cohen, 1988)[2] implies that an SMD of 0.2 is small, 0.5 is medium, and above 0.8 is a large effect size. Considering this, the example difference of 0.53 standard deviations (i.e., 8 points difference in this scale) indicates a medium improvement in receptive vocabulary skills in the intervention group. These standardized versions of the mean difference are particularly important for evidence synthesis, as they allow different studies to be compared on the same scale: standard deviation. For instance, the described study tested improvement in receptive vocabulary with a standardized test which provides z-scores, and the test contains 225 items. However, another study examining effectiveness of reading interventions might have reported a test researchers created, which contains 40 items, and presents a simple summative scoring. Without standardizing these findings, it would be impossible to meaningfully compare their findings. Therefore, evidence synthesis requires each study finding to be transformed into a predefined standardized effect size before combining the results.

Figure 5.5: Description of elements in a forest plot



Mean differences are common effect sizes in studies of educational interventions with group designs. However, there are other possible standardized effect sizes, such as odds ratio or relative risk measures. These effect sizes often describe outcomes in terms of likelihood of the event happening. An educational intervention aimed at improving hand-washing practices of students with intellectual ID could report their findings as odds ratio, describing the likelihood of falling ill after receiving the intervention. Moreso, effect sizes like odds ratio can be used in syntheses of single-case experimental designs, as is demonstrated in Paper II. In Paper II, log

[2]Note that these interpretations are rough guidelines, and depend on the scientific field and phenomena of interest, where these benchmarks might be interpreted differently.

odds were used to represent the increase in the probability of correctly recognizing words after receiving a reading intervention. Selecting an appropriate effect size depends on the type of measure, possible scoring, and distribution of the data. It is not unlikely that measures that seem like they would be captured well with SMDs should in fact be calculated as odds ratio.

Paper II provides reasoning for selecting log odds as an effect size for multiple reasons. First, the synthesis included single-case experimental designs, which focus on individual changes, not on group differences. Second, the interventions in single-case experimental designs often measure behaviours by counting the number of successful trials within one session. For example, in a group design, both groups might receive a 10-item vocabulary test after the intervention period was concluded. Similarly, in single-case experimental designs, students might be tasked with reading 10 items each session. At first glance, the testing instruments seem highly similar. However, in group designs, researchers will often average the summed scores across the groups and compare the means. In single-case designs, these trials are count or proportion measures (in this case bounded – they have a known limit, e.g., 10 items maximum) and behave statistically differently than data from group designs. To put it simply, they can be considered as probabilities of successfully reading words during a session. Furthermore, in these contexts, such as reading interventions, it is expected to see floor and ceiling effects (e.g., the entire baseline phase consisting of no correctly read words), which should be described under binomial or similar distributions. In Paper II, a beta-binomial distribution was used as it accounts for overdispersion of the outcome evident in these extreme values.

Single-case designs are not limited to odds ratio only. There are various other effect sizes existing, and being developed continuously, such as the percentage of non-overlap measures (see Pustejovsky and Ferron, 2017, for an overview). Generally, there are other effect sizes as well, such as correlation coefficients, but for intervention research, standardized mean differences, odds ratio or relative risk are the predominantly used measures.

Meta-analysis. Once effect sizes are calculated, researchers can conduct the meta-analysis. Depending on the model selected, the meta-analysis provides a (average) population effect of the intervention. That is, we get an estimate of how the intervention affects the outcome in the average participant. For example, if the reported meta-analytic effect is 0.5 Hedges' g, that means that on average, the student will improve 0.5 standard deviations after receiving the intervention. Point estimate on its own is not too informative, as we need to be certain that our estimates are precise and consistent. As mentioned in Section 5.2, imprecision and inconsistency should be interpreted together with the point estimate to confirm whether there is evidence for the intervention effectiveness. Figure 5.5 contains an example of a meta-analysis, where the point estimate is $g = 0.3$. On its own, this would be considered a small to moderate effect. The accompanying heterogeneity (τ^2) is 0.02. Not present in the figure, but let us assume the percent of true heterogeneity is 50%. Although moderate, combined with such a small τ^2 lets us estimate the difference between studies as low. However, the confidence interval covers effects to almost 0 (0.03) to 0.57. Therefore, it is unclear whether the true (average) effect size of the intervention is low-to medium, essentially non-existent, or medium. Therefore, it is important to consider all parameters of the effect before considering to believe the effectiveness of the intervention.

Preliminary findings from Paper II exemplify the difference in interpretation when we do not consider the precision of the estimates. For example, the preliminary point estimates of the posterior distribution suggest a large improvement in reading the words upon the introduction of the phonics intervention. Taken in isolation, it would be easy to conclude that phonic interventions are highly beneficial for students with intellectual disability. However, Bayesian methods particularly well describe the uncertainty surrounding the estimates, which allows users to appraise the precision of the modeled parameters. The credible interval estimated along the average estimate of the posterior distribution suggests that the true parameter might even be negative, implying that the intervention might not work. Therefore, it is necessary to incorporate precision estimates when describing the effectiveness of an intervention. Furthermore, we found inconsistency of the effects on the study level. Participant variation was much smaller,

indicating that the leading agents of variation seem to be contextual differences between the synthesized studies. More importantly, Bayesian synthesis additionally estimates precision for heterogeneity parameters. In the preliminary findings from Paper II, we found high uncertainty in heterogeneity estimates as well. Following the guidance of the section explaining how to trust evidence, based on precision and inconsistency estimates, although the point estimates indicate positive effects of the intervention, we cannot confidently conclude that these effects are strong and reliable.

How to evaluate relevance?

Whether an intervention works (i.e., effectiveness and efficacy) is naturally the main interest of both researchers and practitioners. However, the effectiveness is not always universal, and simply because something is found to be effective, does not mean that it is effective always, and for everyone. Pim Cuijpers (Cuijpers, 2026) describes this common misunderstanding in plain language, describing why evidence-based psychotherapies that are on average effective, will not necessarily work on all patients. Essentially, effective interventions that are inaccessible to practitioners for various reasons, or are not appropriate for their students, do not matter much. Therefore, alongside effectiveness, practitioners should evaluate the ratio between effectiveness and cost, and sometimes benefit. Furthermore, an intervention that is evidence-based, does not necessarily always work. Here, I discuss the relevance of evidence-based practices, and characteristics which signal practitioners that the interventions can be feasible to implement in their daily work. To decide on the intervention with the most value and usability, there should be an assessment of its components to identify the amount of required resources, and whether the intervention is applicable in one's specific context. In this section I will provide a guide to identifying key traits of the summarized studies that affect the implementation feasibility, guided by findings from Paper I and Paper II. At the end of this section, I explain which information within evidence syntheses can inform readers about the contexts in which the intervention is applicable.

Cost-effectiveness is a concept linked to the education field by Henry Levin (cf. Levin, 1988; Struening and Guttentag, 1975). Levin introduced the *ingredients method* to estimate costs of interventions. This method involves identifying key ingredients to achieve the effectiveness of the selected intervention, estimating their costs and calculating overall or student-level costs of implementing the intervention. For the purposes of the thesis, I will present the potential characteristics of the study that affect its economic feasibility and general relevance, while the cost estimation is beyond the scope, as it cannot be universally estimated across regions.

For practitioners to be able to implement the intervention, the first relevant information is who implemented the intervention in the studies. Practitioners should be aware whether the implementer has to be specialised in the procedure, have certain degrees, or can just be trained before introducing the practice in their work. Most studies from Paper I had the intervention implemented by the researcher (Table 2 in Paper I). Although some of the researchers had teacher degrees, and often implemented research in their own classroom for the purposes of their research, this additional knowledge that comes from graduate schools might impact the ability to implement the interventions effectively. Some studies had multiple team members involved, either from the research team, the supporting professional staff (e.g., speech and language therapists), and even student peers. Therefore, it is necessary that the evidence synthesis provides information on the person(s) who conducted the interventions. Practitioners should then look for any moderator analyses involving the implementer, or at least be able to make simplified judgments about the feasibility based on the table of characteristics presented in the synthesis. Generally, if an intervention is effective regardless of who implemented the intervention, or if all interventions were conducted by a person that shares practitioner's characteristics, the intervention can be considered in the practitioner's own work.

Secondly, a crucial component of the intervention is the context, i.e., the place where interventions occur. Access to appropriate rooms, with relevant equipment, can make or break an intervention. Paper I only included interventions which were implemented within school settings, but the precise rooms varied across school types and within schools. However, the most commonly used rooms were classrooms. Although these might have been empty classrooms for the purposes of the intervention (which can also affect the intervention), it is important as classrooms are generally accessible to all schools. Following classrooms, many studies had been conducted in various different contexts – those less impactful are probably quiet resource rooms, hallways or offices, as they can most likely be replicated in regular classrooms. There were, however, more specialized rooms for some interventions, such as kitchens or libraries. These might be highly relevant to the success of the intervention, and would most likely not be feasible for regular schools which do not have access to specialized spaces. Therefore, reviews examining effectiveness of interventions should provide information about the context to be able to fully evaluate applicability of the interventions. Context also includes population characteristics. When an evidence synthesis presents findings, it is important to understand for whom the intervention works. Paper I presents the frequencies of diagnoses the participants had. Apart from different or mixed severities of intellectual disability (which were mostly not reported), students with intellectual disability often have co-occurring disabilities. Understanding the extent of the population heterogeneity is crucial for a credible evidence synthesis. Whether an intervention is considered effective on average is of little use to understanding how applicable it is in individual contexts.

Besides context, similarity of interventions and outcomes between synthesized interventions are highly important. First, if they differ, they will introduce heterogeneity. Second, unless all interventions are similar, it is unclear how the practitioner should implement the intervention. Should they use materials from all studies? Should they select those materials and procedures that they find the easiest to implement? How will they know which characteristics and components are actually contributing to the intervention? Overall, if the descriptions of different interventions vary, or if the intervention criteria is not informed by a standardized framework, findings might be difficult or even impossible to implement. Similarly, if the outcomes are operationalized inconsistently, practitioners should be wary of assuming the intervention will be of effect on their outcome of interest, e.g., decoding. If decoding is tested in different ways, practitioners should presume the interventions might not present beneficial results if they are tested differently. Further, longevity and generalizability of the effects are another important feature. Practitioners should look for information about follow-up assessments, their length and number of assessments. Although an intervention might prove to be highly beneficial during the first post-intervention assessment, but the effects might fade away quickly. If that intervention is resource-intensive, it is debatable how useful strong but short-lived effects are. Likewise, generalizability of the effects can be an indicator of how cost-effective the intervention is. For example, sight-word interventions might have more immediate results and larger effects than phonic interventions when it comes to reading trained words, but the generalizability of sight-word intervention effects is uncertain. However, while phonics intervention might intend to develop a universal reading skill, sight-word interventions might be intended to swiftly teach students to read selected words that allows them to safely express their needs. As both interventions serve unique purposes, practitioners should assess their applicability within the context of individuals' needs, not solely based on the evidence strength.

As described above, educational interventions, and particularly interventions in special education, have to handle high heterogeneity. Fisher et al. (2018) discovered the discrepancies between the variance of aggregate values and individuals, and advocate idiographic approaches coupled with open science to remedy human subject research practices. Paper II demonstrates how evidence synthesis can acknowledge individual variation, and utilise data from individual participant data instead of aggregate summaries. This approach allows synthesists to predict effectiveness on a more individualized level. C. R. Cook et al. (2018) call for this personalized approach to education which focuses on *"what intervention worked for whom and how did it*

work?", and meta-analytic approaches similar to the one described in Paper II, which specifically aims to provide those answers. Ultimately, the first step towards unifying incrementalist and reconceptualist approaches will be individualized educational practices, led by the shift towards meaningful engagement with heterogeneity and contextual influences.

Deviations

Unlike in non-preregistered research, there is a high likelihood of finding deviations from protocols that are preregistered. The reason we do not find deviations in non-preregistered research is there is no way to discover them.

While conducting the scoping review, we realized that the scope of identified studies and content expanded our interests beyond the initial objectives related strictly to the characteristics of the design. We recognized meta-scientific patterns, such as inconsistent ethical approval and consent statements, and became interested in the open science practices of the field. Therefore, mid-data extraction, we decided to update our protocol and expand on areas of interest. Additionally, as we identified almost a thousand articles to extract, and found these patterns, we switched from conducting initially planned quality assessment to extracting data about research integrity. This update also included statements about deviations to the protocol which happened to that point, for example switching the software or defining educational interventions more concretely. After the protocol was updated, we proceeded with data extraction.

However, even though we updated the protocol midway, we had to deviate from some sections to keep the project feasible. The report was already significantly larger than most journals allowed, and completion of all segments required more resources than we could afford. For example, in the update, we planned a bibliometric analysis of co-author networks in the field. Furthermore, we only validated 25% of the data extraction, instead of planned double-coding. Although we did not conduct any inferential analyses for this work, and arguably unnecessarily overcommitted with the detailed planning for the type of study, which can be viewed as exploratory, these deviations showcase changes in initial plans and final results, and still provide valuable insights into the process. Had there not been a protocol, and updated changes, the entire process could have been interpreted as much less complex, and the iterativeness would have been hidden.

Similarly, in the process of conducting the expert panel phase for Paper III, there have been updates to the planned study. As the expert panel phase included discussions which resulted in proposals for two versions of the checklist that could be tested in the user panel phase, we used the opportunity to pre-test both versions during a hackathon session at a conference dedicated to metascience and statistical errors, Perspectives on Scientific Error 8. There, we were able to gather initial data that would allow the expert panel to more confidently decide which checklist should be selected for the user panel phase. Although this is technically a part of the planned iterative process between the expert and user phase, it was not explicitly planned as such, and can be considered a deviation. Deviations and protocol updates are unfortunately unavoidable parts of the research process. How little you deviate depends on how thoughtfully you planned the study, and whether there were any external factors that derailed the study (i.e., a pandemic, or major changes in resources and technological accessibility).

When it comes to systematic reviews, deviations that stem from learning about the data are the most severe deviations, and should absolutely be avoided. Deviations concerning the process of obtaining the data might be less consequential. For example, switching software used for screening purposes as it allows for automated deduplication should not affect inferential conclusions, which is an example of a deviation in Paper I. However, realizing that the preregistered model cannot adequately synthesize the study findings can severely undermine the confidence in presented results. Even if the model is corrected, it is effectively a post-hoc analysis, based on seen data, and the severity of the test is now lowered (see Lakens et al., 2024). Consequently,

the severity of these deviations impacts whether a registered report is ultimately published or not. Although deviations are inevitable, there are steps in the systematic review that must be planned as detailed as possible, and piloted to ensure feasibility, to avoid breaching the plan so severely it is deemed irrelevant.

Suggestion Box
Example analysis plan layout per research question

Research Question (RQ) 1
What is the effect of reading on word recognition?

Analysis plan
We will run a meta-analytic multilevel individual-participant data beta-binomial regression to estimate the probability of word recognition, presented as an odds ratio.

Mathematical model
Let

$$\begin{aligned} y_{ijs} &= \text{number of correctly recognised words for participant } j \text{ in study } s, \\ n_{ijs} &= \text{number of words presented,} \\ \text{Reading}_{ijs} &= \text{reading score.} \end{aligned}$$

$$\begin{aligned} y_{ijs} &\sim \text{BetaBinomial}(n_{ijs}, p_{ijs}), \\ \text{logit}(p_{ijs}) &= \beta_0 + \beta_1 \text{Reading}_{ijs} + u_s + u_{js}. \end{aligned}$$

Analytical code

```
library(brms)

model <- brm(
  bf(correct | trials(total) ~ reading +
     (1 | study) + (1 | study:participant)),
  family = beta_binomial(),
  data = dat
)
```

Firstly, unlike scoping reviews, which might not need to be reviewed prior to conduct, systematic reviews should be submitted as registered reports. There are better chances that reviewers will provide additional input in the clarity and feasibility of the plan, and hopefully point out severe issues with the plan before the review process starts. Experience with submitting Paper II to PCI-RR has been extremely helpful in further clarifying our plan, developing the code and providing realistic justifications for our decisions. Especially when developing and using newer methods, receiving input from experts is extremely valuable. It is highly likely we would have run into issues with the analysis had we only preregistered the plan and conducted the study without review. While not bullet-proof, sharing the burden of deciding with external reviewers can help researchers, especially early-career ones, initiate higher quality projects.

Second, instead of only explaining what will be done to answer the question, preregistration should provide detailed explanations of how the question will be answered. This does not include only the generally recommended piloting of search strategies, screening and extraction procedures (recommended by all reputable conduct guidelines, e.g., Cochrane). To avoid deviations, protocols should append the exact code to synthesize the findings in a way that can answer the question. To do this effectively, building a mock dataset which hopefully captures

all expected issues is a valuable practice (missing data, transformations of effect sizes, access to raw data and code to calculate summary statistics, visual data extraction process). With the synthetic data, one can easily play around with different models and test their convergence, possibility to implement, assess run-time, and understand the way the data needs to be organized to streamline the synthesis. It is likely that either the data extraction or models might have to be changed once the synthesis stage approaches, if these details are not properly tested in advance. Furthermore, if the protocol clearly links the textual plan, mathematical model and accompanying code, reviewers can easily gauge feasibility and correctness of the planned analysis. Box 5.2 provides an example of explicitly describing this process, similar to the Study Design table required by *Peer Community In Registered Reports*.

Finally, I suggest keeping the protocol in a form of a living document that can be version controlled. Either through git, OSF files, or other platforms that allow version control (e.g., Google Docs). As you go along, note down where you deviated, why it happened, and how you justify deciding to deviate. In case of deviating severely on a registered report, contact the editor with the updated protocol, clearly showing track-changes in the text. If the work was only publicly preregistered, consider updating the protocol prior to commencing the deviated work, and also explicitly highlight the changes in the subsequent versions. When deviations occur, transparency is key in allowing readers to quickly notice where things changed, and allows them to easily verify the severity of the consequences. A final list of deviations, with clear comparison of planned and conducted research is the simplest way to transparently report deviations. As an example of such reporting, I developed a pre-registration checklist for the journal *Meta-Psychology* (Meta-psychology, 2023) to help authors facilitate the reporting of potential deviations in their submissions.

5.3 Disseminate

Open access is the first step to open science. The first official declaration of support for open access – Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities – was signed 25 years ago (Max Planck Society, n.d.). According to UNESCO's open science initiative (2021), it "means free access to scientific information and unrestricted use of electronic data for everyone." Open access is an umbrella term, and in the field of scientific publishing, it can take several forms. These models of open access usually revolve around types of access to the papers that readers have, and copyright terms for the author and publisher (Barnes, October 22, 2018). The highest level of open access is generally called 'diamond' open access, and provides completely free of charge access to the publication, both for the authors and readers. Additionally, authors typically regain copyright, and the work is published under CC-by[3] license. Typically, journals employ 'hybrid' open access, where the authors can opt for their article being open access by paying additional fees. Copyright licenses can differ depending on the journal, and the article is free to read. Similarly, 'green' open access allows authors to share non-copyedited manuscripts in public repositories, but the journal publication is not free to read. Below I describe how I disseminated the studies in order to follow open access principles.

Journals. Publishing Paper I proved difficult. Surprisingly, the difficult part was finding a relevant, suitable journal, which was supportive of open science practices, encouraged preprinting submitted articles, and most importantly, had a large enough word limit. Literature reviews are long articles, and my methodology section was particularly detailed and long, which most journals did not allow. In turn, I had to move most of my text to supplemental materials, and severely cut down the introduction that justifies the aims of that project. Physically printed articles bring little to no value to the research community nowadays, and word limits serve

[3]Creative Commons licenses are copyright licenses which permit reuse, adaptation and sharing of the materials to certain degrees, depending on level of permissiveness. https://creativecommons.org/share-your-work/cclicenses/

no legitimate purpose, thus these outdated practices still hinder improvement in transparent reporting. Because of this, I could not submit to journals specifically relevant for special education and developmental disabilities. Luckily, I was able to submit my work to an adjacent field of educational research, but these gatekeepers make introducing open science to special education research needlessly challenging. Paper I was published in a hybrid journal, as an open access article, and its accompanying preprint is still available, along with materials openly shared on the OSF.

Preprints. Although the ultimate goal of dissemination in academia is publishing in a peer-reviewed journal, it is often a long process that prevents researchers from disseminating their research to the public. Furthermore, it is costly for the authors to publish in traditional journals, unless they are diamond open access. To combat this, there has been a rise in preprint servers and other dissemination platforms which allow researchers to publicly share their papers as soon as they are written, without cost for the authors or readers. These preprint servers do not audit the quality of the research, and as such do not qualify as a stamp of approval for the shared papers, but they are a way for the articles to reach wider audiences before they are published in journals (Elmore, 2018). All four of my studies have been preprinted, regardless of peer-review and later journal publication process.

Popular Science. Last, but not least, popular science outlets are an important way to reach wider audiences. Blogs, for instance, allow flexibility and control in presenting evidence, facilitating engagement through fun writing styles and less technical diagrams. Writing a blog post about your research provides additional space to point out findings you consider important, and which can be expressed in simpler terms. Blogs are also a great way to post sections of your work and quickly engage with the audience. Beside blogs, social media is an excellent way to disseminate your work among relevant individuals and receive immediate feedback on your work. Popular science venues are incredibly important for fostering scientific literacy and bringing attention to relevant research. It is especially relevant today, when thousands of studies are published daily. Without a strong triage system, studies are most amplified through mainstream media outlets, and researchers can best improve their dissemination skills through engaging with social media and blog platforms. For Paper IV, I shared the preprint on social media. This post was shared and reached members of FORRT (A Framework for Open and Reproducible Research Training) who took great interest in the presented platform. This resulted in an opportunity to write a blog post about the paper, and further share it on FORRT's platform (Batinović & Kalmendal Bruns, 2025). This experience also helped me improve presenting the research to an audience outside of the technical journal systems.

Selecting venues to disseminate research outputs depends on many factors, from monetary to time resources. Unfortunately, what currently drives the peer-review publication system is the aspiration to publish in journals with high impact factor[4]. This drive for journals to obtain high impact factors shifts them towards favoring novel, interesting, but not necessarily high quality or robust studies. Therefore, to reach the full dissemination potential for each publication, and comply with open science practices, the optimal way is to preprint, publish open access, and share in popular science outlets. This approach facilitates early access to feedback, peer-review verification, and a broad outreach to maximise impact. Although this approach might not adhere to the criteria for prestige within the current academic system, it will still allow readers to access the work, while adhering to best practices. Finally, publishing in various outlets encourages exercising the translation of research for different audiences, which in turn improves one's ability to present and clarify different aspects of the research within different contexts.

[4]Impact Factor (IF) is a measure used to evaluate the importance of an academic journal by calculating the average number of citations received by its articles over a certain number of years. Higher IF is associated with higher prestige of the journal.

5.4 Supplement

Supplemental information is the essence of the Open Evidence Synthesis Workflow. Although mandated by large evidence synthesis organizations (e.g., Cochrane and Campbell), sharing materials which facilitate reproducibility is still not the norm outside that realm. In fact, a recent large-scale attempt to reproduce and replicate studies discovered the poorest reproducibility and open science practices in educational research (Miske et al., 2026). This is a concerning finding, and to an extent, inexcusable.

Main components that should be shared are the protocol, search strategies, extracted data and accompanying documentation, and the analysis. Sharing materials is what makes the synthesis open, and ultimately reproducible. However, to be *fair*, providing supplements in a FAIR (Findable, Accessible, Interoperable, Reusable) format is challenging. The FAIR standards propose dissemination of research data in a findable, accessible, interoperable, and reusable manner. In this section, I describe the process, challenges, and recommendations for complying to FAIR standards, based on the experience of conducting the four studies. The concrete descriptions of repositories I used can be considered illustrative, as these repositories go through changes and updates and will most likely be outdated by the time this thesis goes public. Principles matter more, and exact tools used to follow FAIR guidelines should be based on individual needs of the researcher.

Data

Data is the core outcome of a study. All scientific findings are based on data collected in some form. To be able to verify those findings, access to that data is the necessary first step. Therefore, data sharing is the pillar of credible science. However, for data to be useful, it should be shared in a legible format, with accompanying descriptions that facilitate reproduction. Formatting the data in a shareable format is a time-consuming process. It is especially complex when the data contains sensitive, identifying personal information, and requires anonymization. For the scoping and systematic reviews, that was not an issue. For the Delphi Panel, personal data of the participants was removed, and items from each answer sheet were collated into one. This way individual participant answers were available, but the risk of tracking down individual participants was minimized.

Cleaning and preparing the data is necessary to comply to the *Reusable* principle of the FAIR framework. The cleaning process should ensure the dataset is in a format that can be manipulated by software that is generally accessible (preferably free or open source) and complies to the standards of the scientific field aimed at reusing them. In my case, I shared my data files as .ods and .xlsx files. .ods ensured that users without access to proprietary softwares can open and use the file, while the .xlsx format was shared because it is closer to Google Sheets in functionality and would ensure no (or minimal) data loss occurs when saving it in that format. Generally, the preferred formats for archiving tabular data are the simplest file formats, such as .txt, or .csv. However, formats like .xlsx are convenient for further editing and richer in data, and potentially functions, so as long as they are available in open source formats (e.g., .ods), the files are still reusable. *Accessibility* simply refers to the ability of accessing the data solely with a computer and internet connection. Although it is debatable whether that is true accessibility, it is hard to imagine that access to data in other formats would be of use regardless, in case these conditions are not met. For example, the OSF is a free platform, available anywhere in the world (where you have access to a computer and internet), and the data and metadata can be accessed by simply clicking on the link leading to it. To ensure *Findability* of the data, I shared the datasets on the Open Science Framework (OSF). Following their guidelines, the datasets are archived in a separate component, filed as "datasets", with metadata describing the purpose and content of the component. The component has its own persistent identifier (e.g., DOI), which is linked in the related publications. DOI and metadata are crucial for findability, and repositories such as OSF and Zenodo already integrate an interface that facilitates metadata creation. The key is to use

that interface and provide necessary information for the metadata to be useful. Lastly, *Interoperability* is achieved by properly linking all related (meta)data to each other. Perhaps the simplest way to achieve this through OSF is by uploading all connected datasets into one component. Although it is recommended that each component represents one dataset, this seems wasteful and overly complex, which most educational researchers will not do, nor will concretely help potential researchers looking to reuse their data.

Meta-researchers, and the scientific community in general will eventually benefit from following the FAIR standards, however, from personal experience of following these standards, they genuinely are resource demanding, with little immediate tangible benefit to the researcher. Reusing data is still not the norm, and researchers still tend to be introduced to data after reading the accompanying articles, rather than scrolling through data repositories. For those purposes, and the purposes of research synthesists, providing DOIs to a dataset is much more beneficial than focusing on metadata enrichment. Understanding that open science requires additional work is not however and excuse for researchers to avoid sharing data. But for now, focusing on the data being accessible and reusable would be impactful enough, without worrying about optimizing the findability (which is expected to naturally come with the written report) or interoperability (which can simply be explained in the report or a README file).

Materials

Unlike data, materials are crucial for reproducing and replicating the study. Reproducing refers to rerunning the same analysis and procedure on the same data to get identical results. Replicability refers to conducting the same procedure to collect new data. Replicability of evidence synthesis is as important as replicability of original research. Sharing materials is also possible under the FAIR framework. All components follow the same principles as when sharing data, with particular emphasis on reusability.

The most important components to share to replicate an evidence synthesis are the protocol, instructions for screening and data extraction, and the search strategies. To further reproducibility of the work, supplementing the searches with files including initially identified, and screened references ensures the entire evidence synthesis procedure can be reproduced, beyond simply replicating the study on new data (i.e., additional or modified search results). Reusability is achieved by sharing complete search strings and details about each searched datatabase, which can preferably be edited, or copy-pasted without additional manual modifications. This implies sharing the searches in files such as textual files or Microsoft Word document. Portable document files (PDF) are not suitable in this case. Searches can even be disseminated on servers dedicated to hosting review search strategies, i.e., *searchRxiv*.

I generally gravitate towards the repositories I used to share materials to look for my projects because those projects contain essential and cleaned up documents, which also means I get to experience what I have shared from others' point of view. This also helps me with future projects that rely on my previous materials, and it is of great use that I can download and simply continue editing my protocol and search strategy files. On the other hand, I have shared most screening guides in non-editable PDFs, which in hindsight is not the best option for reusability, although it is a more accessible format. However, I do not think it is cost-effective to provide multiple formats of the same file just to ensure both human- and machine-readability, which is more relevant for data sharing.

Analysis

Reproducibility of the study's findings depends on sharing data and the analysis files. For a fully reproducible analysis package, users should have access to a README file which explains all components of the package, a codebook that explains each variable in the dataset, analysis scripts and output in an easily reproducible format. This applies to analyses from original

studies and evidence syntheses. Hornung et al. (2026) shared a checklist of recommended files researchers should provide to facilitate reproducibility.

Ensuring reproducibility is not easy. Even with proper documentation, and sharing all files, re-running analyses requires machine compatibility and often is not as straightforward as it seems. For example, I have shared my analysis files online for Paper II, which involved more complex Bayesian analyses. They usually require longer computing time and several programming packages. This resulted in one of the reviewer being unable to set up and run my scripts. To circumvent that, I uploaded the scripts to CodeOcean, an online container emulator, to ensure anyone can re-run the analyses without having to locally run them. Sharing the analyses is not only useful so others can confirm the accuracy of your work, but to reuse it and build on it. Many of the studies I have read with the aim of implementing their proposed methods and conducted analyses, I could not implement because they did not provide code for the analyses they conducted. Granted, not all researchers are comfortable with sharing the code that produced the analyses, especially in social sciences where high-level coding is not the core skill. However, providing the code ensures that the mathematical operations and code are compatible and the analysis is conducted the way it was planned to be conducted.

Most importantly, as long as the code is properly commented and accompanied with documentation that explains the hierarchy between files, most users can verify the findings, and potentially reuse the materials. Although proper documentation is a tedious, and also complex task, it is one of the places where LLMs excel, and can help researchers write high-quality documentation quickly – naturally, with constant human supervision and final confirmation.

5.5 Maintain

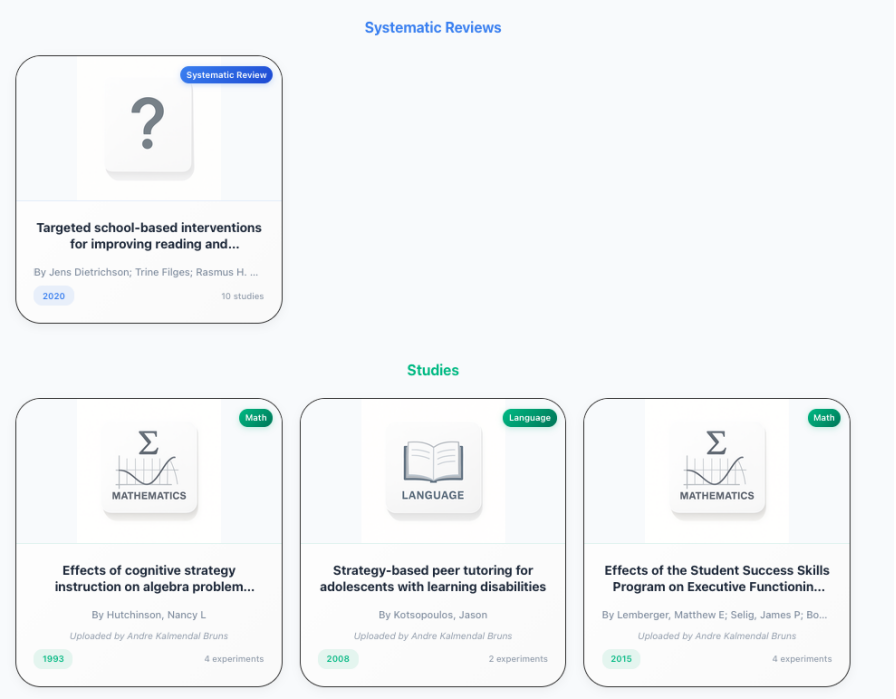
Paper IV: Make it dynamic

The final paper presented in this thesis is an introduction to an open-source platform that allows users to conduct their own meta-analyses by selecting intervention studies from a continually updated living database. This platform is the end goal of providing evidence-based practices in a format that is accessible to researchers, as well as practitioners.

Regular evidence syntheses are generally static, meaning that the presented evidence is a snapshot of studies included up to the point of publication. Inflexibility of published reviews can be beneficial to a well-researched phenomenon that has a large, high-quality body of evidence and non-evolving characteristics. In that case, the static review provides a definitive conclusion. Static reviews still do not enable readers to conduct moderator analyses of their interest, but this is less concerning. However, for newly developing evidence (e.g., new AI based interventions), static reviews cannot follow the evidence and are often old before they even get published. For these instances, a different type of systematic reviews is recommended – a living systematic review (Elliott et al., 2017). Living systematic reviews follow the same procedure as regular systematic review, with additional input on search frequency and plan for integrating new evidence. Search frequency is much higher than regular expectations of updating systematic reviews (Cochrane recommends within 5 years), and could be done as often as weekly, as was the case during the COVID-19 pandemic. Most often, search frequencies depend on the frequency of database updates, or can be done as soon as new evidence is flagged by reviewers or any involved third party (e.g., web-based notification systems, readers). Then, the evidence is assessed and integrated based on its implications for the conclusions. Although living reviews are more timely than regular systematic reviews, bottlenecks still occur. New research has to be identified, extracted by the reviewers, assessed for importance, and updated, which then leads to waiting for publication.

To resolve these issues around sustainability and feasibility, dynamic and crowd-sourced solutions have been suggested (Thomas et al., 2017; Tsuji et al., 2014). Community-augmented meta-

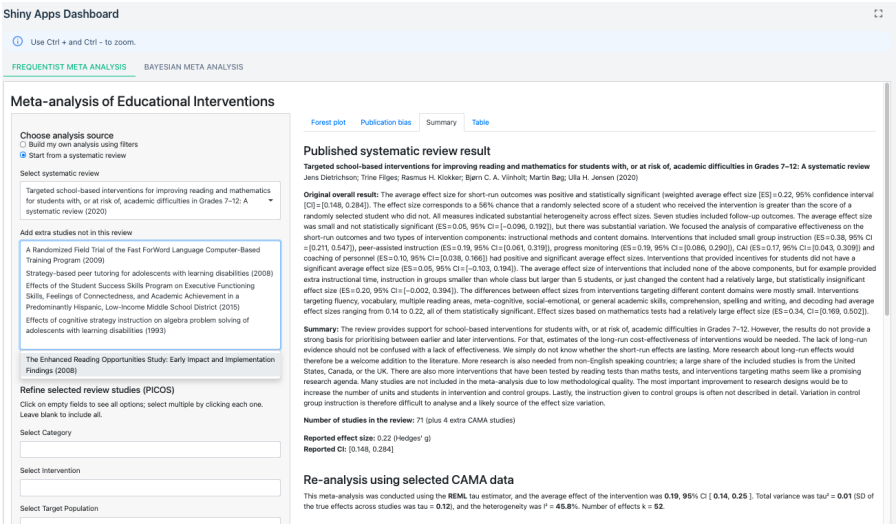
Figure 5.6: Display of individual studies and systematic reviews on the platform



analysis (CAMA) is one of the earliest attempts to provide dynamic access to meta-analytic evidence which is regularly updated through crowd-sourced efforts. Following these principles, I present an open-source platform which allows evidence synthesis of school-based interventions identified in the scoping review. Although Tsuji et al. (2014) provided code for their CAMA platform, in the process of creating our own version, my collaborators and I shifted towards a more complex system that included a separate database of interventions, and a more comprehensive landing page that provides insight into the characteristics of the studies. All components of the platform used to build it and platform itself are open-source, i.e., the code is freely available for reuse, modification and redistribution. Paper IV introduces the platform and provides a use-case example.

The database was built in SQL (Structured Query Language), a programming language specialized for development, querying and maintaining relational databases. More specifically, we used PostgreSQL (postgresql.org/), which is an open-source SQL system. Because of SQL, the entire platform and the dashboard can easily access and retrieve parts of the database depending on interest, or access all of the data. This approach allows scalability and speed of information retrieval, which makes the platform usable for multiple simultaneous users. The front-end, i.e., the website section that the users interact with, was built in Django (djangoproject.com), a Python-based framework, selected as it is also open-source. Meta-analyses are available through two dashboards, providing users with a frequentist and a Bayesian alternative for the analyses. Both dashboards were built with RShiny, an R package created for web development programmed in R programming language (R Core Team, 2024). The platform enables continuous updating of research evidence from educational interventions. For example,

Figure 5.7: Example summary tab from the CAMA platform



the results from the scoping review provide the foundational set of eligible studies, which can be meta-analysed and whose quality will have been assessed beforehand. Once the platform is established, populating with new studies can continue with regular updating of searches created for the reviews (Paper I) or by community involvement of researchers who find eligible studies and upload the data to the platform. Figure 5.6 provides an example of how individual studies are displayed on the platform.

One of the primary goals of the platform is to provide a summary of results from meta-analyses in a format that is more accessible to practitioners. With that in mind, the platform currently provides a simple boilerplate summary of the results and a simple description of the practical meaning behind the numbers. Furthermore, users can meta-analyze collections of studies extracted from a published systematic review, and modify the analysis with additional studies not included in the review. The summary tab presents the published results and the new results from the user's analysis (see Figure 5.7). However, to establish summaries that fit each meta-analysis in more detail, there needs to be more research into meaningful interpretation of effect sizes, and cost-effectiveness assessments of the displayed interventions.

Overall, the platform enables augmenting the research that led to its development, i.e., hosts data collected through the scoping and systematic reviews, with evaluated quality and research integrity. This stage completes the workflow but ensures the evidence is continually updated and represents the most recent state of the art. As the platform gets populated, it will serve as a hub for furthering research in a responsible manner, by conducting research where gaps are identified.

6 Implications

While reviewing the intellectual disability research (Paper I), I have come across various ways of defining the constructs that were measured in the identified studies. I became increasingly aware of the diversity in operational definitions, particularly regarding the population of interest. This variance can perhaps be explained through the relativistic approach, as discussed by Benton and Craib (2011), which provides a framework for understanding this variety of definitions. Furthermore, the criteria was interpreted loosely and often pragmatically, as not to lose participants over overly restrictive inclusion criteria. This approach to research in special education is unavoidable, and to some extent promotes equity within research, as promoted by pluralists (Baglieri et al., 2011). However, the diversity of diagnoses, paired with subpar reporting practices and lack of data sharing poses serious threat to credible evidence synthesis, and contributes to research waste.

Researchers who plan systematic reviews with a focus on students with intellectual disability should carefully curate search strategies and consult experts that will ensure they capture all relevant terminology. Furthermore, I strongly recommend spending additional time planning concrete objectives for extraction of the data necessary for a meta-analysis to ensure consistency and homogeneity of designs. Meta-analytic studies are often haunted by the "garbage-in-garbage-out[1]" monster, and the subpar reporting practices could exacerbate inappropriate synthesis of interventions. Unfortunately older research cannot be ammended by improved reporting standards, but it is an urgent time to start sharing research materials and data. Evidence synthesis is a fragile method, and errors are more likely when data is poorly described and presented only in summary statistics. Researchers conducting original research will benefit the overall evidence base by following reporting standards and sharing materials. Particularly with the

[1]Garbage-in, garbage-out (GIGO) is a colloquial term, stemming from computer science, which describes the consequences of analyzing poorly entered data. Particularly relevant for meta-analysis, it implies that the meta-analytic results can only be as good as the quality of included studies and the quality of handling those studies.

today's availability of online repositories, there are no challenges with quality reporting due to space-limited journal articles. Furthermore, making data and materials available will augment replications by allowing conceptually and methodologically more similar interventions.

Besides the conceptual diversity, special education research employs a variety of research designs. This calls for further developments in meta-analytic methods. Paper II is a case-example of adapting meta-analytic models to utilize data provided by single-case designs. There is a movement towards establishing new statistical methods to best analyze these methods, and move away from the medicine-imposed standards centered around randomized controlled trials. If the field does not adapt to the practices better suited for low-incidence populations, risk of research waste is high. Although adaptations to accommodate different study designs is necessary, it is not without its own challenges. Randomized controlled trials are considered gold-standard for a reason. As mentioned previously, evidence synthesis is a fragile method, susceptible to bias on multiple levels. Synthesizing solely RCTs alleviates many of possible causes of bias. However, quasi-experimental group and single-case designs are not without worth, and can be strong designs when done well. Nevertheless, the study design itself is no longer the only threat to evidence integrity. Research integrity assessment has become a necessary step in systematic reviews, as evaluation of risk of bias in untrustworthy studies is a waste of time at best, and misleading at worst. It is imperative that researchers conducting evidence synthesis start implementing forensic metascience in the review workflow. Development of the research integrity checklist presented in Paper III is only a small step towards systematizing this type of research evaluation, and there is a high need for research synthesists to develop forensic metasience skills to properly keep up with the changes in the scientific landscape.

Finally, as a meta implication of the work presented in this thesis, with the disclaimer that I have mostly relied on my own data, I found that open science practices genuinely expedite scientific endeavor. I have re-used shared extractions from Paper I to conduct the bulk of the data extraction in Paper II. Although not everything relevant for Paper II was available in the dataset from Paper I, the ground work conducted previously facilitated building on top of it. I have also greatly benefited from researchers who have made their code available (e.g., Grekov et al., 2025) to build my own analyses. Furthermore, by participating in peer review of the presented IPA protocols, studies were evaluated for their accuracy and realism before we committed resources into their realization. With the data and search strings from Paper I shared online, anyone can select a relevant domain and do another evidence synthesis similar to Paper II, by reusing the majority of materials. Regardless of inevitable modifications, building on an existing foundation is always simpler. Every step of the research process was made easier for me because my predecessors engaged in open science practices. Open science was not only embedded into the conduct of this dissertation, it made it possible.

6.1 Future directions

Findings presented in this dissertation have barely scratched the surface. Technological developments are advancing daily, while the intellectual disability education research field struggles to keep up with the changes. To combat the malevolent use of new advancements in research, this field needs to establish clear standards for producing strong, open research that will guide evidence-based practices.

To accomplish this goal, there is a high need to continue meta-research in the area of special education. Specifically, systematic assessments of adherence to reporting standards and scientometric assessment of production in this area would guide improvement of reporting transparency and facilitate reproducibility. It is crucial to understand the current reporting practices and the systemic failures to report necessary information. As there are already strides towards accessible and transparent and research in other fields within intellectual disability research (Laverty & Richards, 2026), special education research has the tools to follow suit. Additionally, scientomet-

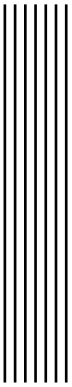
ric evaluations would augment understanding which publications are more likely, and whether any scientific practices are associated with the incentives from the publication system, rather than the pursuit of knowledge.

Paper IV illustrated one of the approaches to living reviews and the utilization of existing evidence. However, dynamic synthesis platforms are heavily geared towards accommodating group designs with established synthesis methods. This dissertation illuminated the diverse and large corpus of single-case designs which are currently omitted from the conversations about living evidence synthesis. As showcased in Paper II, there is untapped potential in addressing individual variations and synthesizing single-case designs. Similar methodological advances are happening elsewhere, with an uptake of evidence synthesis of single-case designs (e.g., Kok et al., 2026). Therefore, efforts should be put towards developing dynamic platforms that can accommodate single-case design research and augment the existing evidence base. Furthermore, research assessing the spectrum of effect sizes common in the field would lead to establishing meaningful benchmarks, while assessment of terminology and the use of instruments and diagnostic tools would help alleviate heterogeneity concerns (Elson et al., 2023; Topor & Danielsson, 2026). Finally, future research should focus on evaluating the the formalization of research questions, appropriateness of operationalized constructs, and utilization of theories in development of interventions. To this end, the field should move towards multi-lab, big team collaborations, particularly aiming for harmonized attempts to conduct interventions for a prospective meta-analysis. With particular emphasis on developing theory-guided interventions, large scale trials are the most efficient way to establish evidence-based practices.

7 Conclusion

The current landscape of intervention research for students with intellectual disability highlights the need to continue advocating for open science practices for the advancement of evidence-based practices. The aim of this dissertation was to gain an understanding of the factors that affect the quality of evidence synthesis in special education research and develop methods to effectively synthesize available evidence, thus improving the quality of evidence synthesis. Namely, I identified general research patterns through a comprehensive scoping review, which indicated areas with potential for methodological developments (Paper I). The discovered diverse set of intervention designs, particularly single-case experimental designs, suggested an underutilized potential for development of new synthesis methods. Paper II was a direct attempt to synthesize single-case studies while following open science practices. Preliminary findings present new challenges and opportunities for evidence synthesis of single-case research. The traditional graphical visualization of raw data provided an opportunity for an individual-participant meta-analysis. Despite accompanied subpar reporting standards, we maximised the use of shared data. By implementing Bayesian methods, we were able to synthesize small samples inherent to this research and learn about the predicted probabilities of reading after completing phonics interventions. Furthermore, we were able to learn about heterogeneity on study and participant levels, and better understand where variation originates. In our preliminary findings, the heterogeneity stemmed from the studies indicating important contextual factors that could moderate effectiveness of reading interventions, and better reporting standards of methods and materials would facilitate identification of these contextual agents. Patterns identified through Paper I indicated that the research integrity of primary studies needed to be systematically addressed and evaluated. These findings prompted the development of a research integrity checklist aimed at facilitating screening of untrustworthy papers in evidence syntheses, which was presented in Paper III. Together, these studies provided a foundation to filter trustworthy evidence and prevent research waste by synthesizing research designs common in the field of intellectual disability education, all to present evidence in a dynamic, accessible matter on the research synthesis platform developed as part of the dissertation. The platform, presented in Paper IV, was developed to support living evidence synthesis, allowing continuous updating of

the evidence base which would ultimately inform evidence-based practices. Through the findings presented in this thesis, we learned that the absence of open science practices and research integrity concerns affect the quality of evidence synthesis. Further, it was discovered that evidence base highly relies on undersynthesized research designs. To improve the quality of open evidence synthesis, the studies in the thesis were pre-registered and shared materials and data, focused on openly synthesizing available evidence and developed research integrity assessment methods to increase the trustworthiness of evidence syntheses in special education.



Bibliography

Abalkina, A., & Bishop, D. (2023). Paper mills: A novel form of publishing malpractice affecting psychology. *Meta-Psychology*, *7*. https://doi.org/10.15626/MP.2022.3422

Altman, D. G. (1994). The scandal of poor medical research. *BMJ*, *308*(6924), 283–284. https://doi.org/10.1136/bmj.308.6924.283

American Association of Intellectual and Developmental Disabilities (AAIDD). (2010). *Intellectual disability: Definition, classification, and systems of supports, 11th ed. of AAIDD definition manual*. American Association on Intellectual; Developmental Disabilities.

American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders, fifth edition, text revision (DSM-5-TR)*. https://www.appi.org/dsm5tr

Amir-Behghadami, M., & Janati, A. (2020). Population, intervention, comparison, outcomes and study (PICOS) design as a framework to formulate eligibility criteria in systematic reviews. *Emergency Medicine Journal*, *37*(6), 387. https://doi.org/10.1136/emermed-2020-209567

Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, *8*(1), 19–32. https://doi.org/10.1080/1364557032000119616

Aromataris, E., Lockwood, C., Porritt, K., Pilla, B., & Jordan, Z. (Eds.). (2024). *JBI Manual for Evidence Synthesis*. JBI. https://doi.org/10.46658/JBIMES-24-01

Baglieri, S., Valle, J. W., Connor, D. J., & Gallagher, D. J. (2011). Disability Studies in Education: The Need for a Plurality of Perspectives on

Disability. *Remedial and Special Education*, *32*(4), 267–278. https://doi.org/10.1177/07419325103622000

Barnes, L. (October 22, 2018, October). Green, Gold, Diamond, Black – what does it all mean? https://doi.org/10.11647/OBP.0173.0089

Batinović, L., & Kalmendal Bruns, A. (2025, April). Community-Augmented Meta-Analyses in Education (ELD CAMA) | FORRT. Retrieved September 15, 2026, from https://forrt.org/educators-corner/020-community-augmented-meta-analyses-in-education/

Batinovic, L., Topor, M., & Pickering, J. (2025, February). Gold in, gold out. Quality appraisal and risk of bias tools to assess non-intervention studies for systematic reviews in the behavioural sciences: A scoping review [Stage 1 IPA at PCI RR]. Retrieved January 4, 2026, from https://osf.io/4gy5b

Benton, T., & Craib, I. (2011). *Philosophy of social science: The philosophical foundations of social thought* (2. ed., 10th anniversary ed). Palgrave Macmillan.

Biggs, E. E., & Robison, S. E. (2023). Review of the Evidence Base for Peer Network Interventions for Students With Intellectual and Developmental Disabilities. *Remedial and Special Education*, *44*(1), 43–59. https://doi.org/10.1177/07419325221075013

Bouck, E. C. (2012). Secondary students with moderate/severe intellectual disability: Considerations of curriculum and post-school outcomes from the National Longitudinal Transition Study-2. *Journal of Intellectual Disability Research*, *56*(12), 1175–1186. https://doi.org/10.1111/j.1365-2788.2011.01517.x

Bouck, E. C. (2014). The postschool outcomes of students with mild intellectual disability: Does it get better with time? *Journal of Intellectual Disability Research*, *58*(6), 534–548. https://doi.org/10.1111/jir.12051

Bowen, A. E. J., Ferreira, R. A., Tolmie, A., Thomas, M. S. C., Borst, G., & van Herwegen, J. (2025). International perspectives on gaps and solutions for integrating research evidence into classroom practices. *npj Science of Learning*, *10*(1), 79. https://doi.org/10.1038/s41539-025-00370-x

Bowman-Perrott, L., Gilson, C., Boon, R. T., & Ingles, K. E. (2023). Peer-Mediated Interventions for Students with Intellectual and Developmental Disabilities: A Systematic Review of Reviews of Social and Behavioral Outcomes. *Developmental Neurorehabilitation*, *26*(2), 134–154. https://doi.org/10.1080/17518423.2023.2169878

Buckley, J., Hyland, T., & Seery, N. (2023). Estimating the replicability of technology education research. *International Journal of Technology and Design Education*, *33*(4), 1243–1264. https://doi.org/10.1007/s10798-022-09787-6

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*, 1–28. https://doi.org/10.18637/jss.v080.i01

Burns, M. K., & Ysseldyke, J. E. (2009). Reported Prevalence of Evidence-Based Instructional Practices in Special Education. *The Journal of Special Education*, *43*(1), 3–11. https://doi.org/10.1177/0022466908315563

Carlsson, R., Danielsson, H., Heene, M., Innes-Ker, Å., Lakens, D., Schimmack, U., Schönbrodt, F. D., van Asssen, M., & Weinstein, Y. (2017). Inaugural Editorial of Meta-Psychology. *Meta-Psychology*, *1*. https://doi.org/10.15626/MP2017.1001

Carnine, D. (1999). Campaigns for Moving Research Into Practice. *Remedial and Special Education*, *20*(1), 2–35. https://doi.org/10.1177/074193259902000101

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1), 1–32. https://doi.org/10.18637/jss.v076.i01

Carter, M., Stephenson, J., & Strnadová, I. (2011). Reported prevalence by Australian special educators of evidence-based instructional practices. *Australasian Journal of Special Education*, *35*(1), 47–60. https://doi.org/10.1375/ajse.35.1.47

Chalmers, I., & Glasziou, P. (2009). Avoidable waste in the production and reporting of research evidence. *The Lancet*, *374*(9683), 86–89. https://doi.org/10.1016/S0140-6736(09)60329-9

Chalmers, I., Hedges, L. V., & Cooper, H. (2002). A Brief History of Research Synthesis. *Evaluation & the Health Professions*, *25*(1), 12–37. https://doi.org/10.1177/0163278702025001003

Chambers, C. D., & Tzavella, L. (2022). The past, present and future of Registered Reports. *Nature Human Behaviour*, *6*(1), 29–42. https://doi.org/10.1038/s41562-021-01193-7

Chen, M., & Pustejovsky, J. E. (2024). Multilevel meta-analysis of single-case experimental designs using robust variance estimation. *Psychological Methods*, *29*(3), 537–560. https://doi.org/10.1037/met0000510

Chen, M., Pustejovsky, J. E., Klingbeil, D. A., & van Norman, E. R. (2023). Between-case standardized mean differences: Flexible methods for single-case designs. *Journal of School Psychology*, *98*, 16–38. https://doi.org/10.1016/j.jsp.2023.02.002

Clayton, M. J. (1997). Delphi: A technique to harness expert opinion for critical decision-making tasks in education. *Educational Psychology*, *17*(4), 373–386. https://doi.org/10.1080/0144341970170401

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

Collins, L. W., Cook, S. E. C., & Cook, B. G. (2025). Evidence-Based Practices in Special Education: A Brief Evolution and Noteworthy Practices. *TEACHING Exceptional Children*, 76–77. https://doi.org/10.1177/00400599251382310

Cook, B. G., Buysse, V., Klingner, J., Landrum, T. J., McWilliam, R., Tankersley, M., & Test, D. W. (2015). CEC's Standards for Classifying the Evidence Base of Practices in Special Education. *Remedial and Special Education*, *36*(4), 220–234. https://doi.org/10.1177/0741932514557271

Cook, B. G., & Cook, S. C. (2013). Unraveling Evidence-Based Practices in Special Education. *The Journal of Special Education*, *47*(2), 71–82. https://doi.org/10.1177/0022466911420877

Cook, B. G., Fleming, J. I., Hart, S. A., Lane, K. L., Therrien, W. J., van Dijk, W., & Wilson, S. E. (2022). A How-To Guide for Open-Science Practices in Special Education Research. *Remedial and Special Education*, *43*(4), 270–280. https://doi.org/10.1177/07419325211019100

Cook, B. G., Therrien, W. J., Waterfield, D. A., McClain, S., Fleming, J. I., Robinson, H., Watson, L., & Boyle, J. (2025). Same as It Ever Was: An Updated Review of Replication Studies in Special Education Journals. *Remedial and Special Education*, *46*(3), 236–248. https://doi.org/10.1177/07419325241248766

Cook, C. R., Kilgus, S. P., & Burns, M. K. (2018). Advancing the science and practice of precision education to enhance student outcomes. *Journal of School Psychology*, *66*, 4–10. https://doi.org/10.1016/j.jsp.2017.11.004

Cooke, A., Smith, D., & Booth, A. (2012). Beyond PICO: The SPIDER tool for qualitative evidence synthesis. *Qualitative Health Research*, *22*(10), 1435–1443. https://doi.org/10.1177/1049732312452938

Costanzo, F., Varuzza, C., Menghini, D., Addona, F., Gianesini, T., & Vicari, S. (2013). Executive functions in intellectual disabilities: A comparison between Williams syndrome and Down syndrome. *Research in Developmental Disabilities*, *34*(5), 1770–1780. https://doi.org/10.1016/j.ridd.2013.01.024

Cuijpers, P. (2026, July). Why does a treatment not work, while it is evidence-based? Retrieved July 11, 2026, from https://www.metapsy.org/patients/evidence-based-does-not-work.html

Cumpston, M., Flemyng, E., Thomas, J., Higgins, J. P. T., Deeks, J. J., & Clarke, M. J. (2023). Chapter I: Introduction. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane handbook for systematic reviews of interventions* (Version 6.4). Cochrane. https://training.cochrane.org/handbook

Danielsson, H., Henry, L., Messer, D., & Rönnberg, J. (2012). Strengths and weaknesses in executive functioning in children with intellectual disability. *Research in Developmental Disabilities*, *33*(2), 600–607. https://doi.org/10.1016/j.ridd.2011.11.004

Danielsson, H., Zottarel, V., Palmqvist, L., & Lanfranchi, S. (2015). The effectiveness of working memory training with individuals with intellectual disabilities - a meta-analytic review. *Frontiers in Psychology*, *6*(1230). https://doi.org/10.3389/fpsyg.2015.01230

Day, E., Steinka-Fry, K., Schweer-Collins, M., Grant, S., Trevino, S., & Tanner-Smith, E. E. (2025). Talking at cross purposes? Exploring (mis)alignment between research and practice. *Educational Research*, *67*(3), 265–285. https://doi.org/10.1080/00131881.2025.2529256

Denny, L., Coles, S., & Blitz, R. (2017). Fetal Alcohol Syndrome and Fetal Alcohol Spectrum Disorders. *American Family Physician*, *96*(8), 515–522.

Diamond, I. R., Grant, R. C., Feldman, B. M., Pencharz, P. B., Ling, S. C., Moore, A. M., & Wales, P. W. (2014). Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. *Journal of Clinical Epidemiology*, *67*(4), 401–409. https://doi.org/10.1016/j.jclinepi.2013.12.002

Eisenman, L. T. (2007). Self-determination interventions: Building a foundation for school completion. *Remedial and Special Education*, *28*(1), 2–8. https://doi.org/10.1177/07419325070280010101

Eiser Hess, A., Cook, B., Pfeiffer, D. L., Ford, A. L. B., & Reed, A. C. (2025, May). *Open-Science Practices in Two Communication Sciences and Disorders Journals: A Systematic Review* [Preprint]. https://doi.org/10.35542/osf.io/6habn_v1

Elliott, J. H., Synnot, A., Turner, T., Simmonds, M., Akl, E. A., McDonald, S., Salanti, G., Meerpohl, J., MacLehose, H., Hilton, J., Tovey, D., Shemilt, I., Thomas, J., & Living Systematic Review Network. (2017). Living systematic review: 1. Introduction-the why, what, when, and how. *Journal of Clinical Epidemiology*, *91*, 23–30. https://doi.org/10.1016/j.jclinepi.2017.08.010

Elmore, S. A. (2018). Preprints: What Role Do These Have in Communicating Scientific Results? *Toxicologic pathology*, *46*(4), 364–365. https://doi.org/10.1177/0192623318767322

Elson, M., Hussey, I., Alsalti, T., & Arslan, R. C. (2023). Psychological measures aren't toothbrushes. *Communications Psychology*, *1*, 25. https://doi.org/10.1038/s44271-023-00026-9

Engel, G. L. (1977). The Need for a New Medical Model: A Challenge for Biomedicine. *Science*, *196*(4286), 129–136. https://doi.org/10.1126/science.847460

Faust, D. (2005). Why paul meehl will revolutionize the philosophy of science and why it should matter to psychologists. *Journal of Clinical Psychology*, *61*(10), 1355–1366. https://doi.org/10.1002/jclp.20185

Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research.

Proceedings of the National Academy of Sciences, *115*(27), E6106–E6115. https://doi.org/10.1073/pnas.1711978115

Fleming, J. I., McLucas, A. S., & Cook, B. G. (2023). Review of Four Preregistration Registries for Special Education Researchers. *Remedial and Special Education*, *44*(6), 495–505. https://doi.org/10.1177/07419325231160293

Fleming, J. I., Wilson, S. E., Espinas, D., van Dijk, W., & Cook, B. G. (2025). Special Education Researchers' Knowledge, Attitudes, and Reported Use of Open Practices. *Remedial and Special Education*, *46*(5), 359–370. https://doi.org/10.1177/07419325241237268

French, A. (2022, November). Case Study: ROR in the COS Open Science Framework [Medium: text/html]. https://doi.org/10.71938/GDR4-D136

Gage, N. A., Cook, B. G., & Reichow, B. (2017). Publication bias in special education meta-analyses. *Exceptional Children*, *83*(4), 428–445. https://doi.org/10.1177/0014402917691016

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis* (3rd ed.). Chapman & Hall/CRC.

Gersten, R., Fuchs, L. S., Compton, D. L., Coyne, M. D., Greenwood, C. R., & Innocenti, M. S. (2005). Quality Indicators for Group Experimental and Quasi-Experimental Research in Special Education. *Exceptional Children*, *71*(2), 149–164. https://doi.org/10.1177/001440290507100202

Glass, G. V. (1976). Primary, Secondary, and Meta-Analysis of Research. *Educational Researcher*, *5*(10), 3–8. https://doi.org/10.2307/1174772

Goering, S. (2015). Rethinking disability: The social model of disability and chronic disease. *Current Reviews in Musculoskeletal Medicine*, *8*(2), 134–138. https://doi.org/10.1007/s12178-015-9273-z

Goodman, C. M. (1987). The Delphi technique: A critique. *Journal of Advanced Nursing*, *12*(6), 729–734. https://doi.org/10.1111/j.1365-2648.1987.tb01376.x

Grainger, M. J., Bolam, F. C., Stewart, G. B., & Nilsen, E. B. (2020). Evidence synthesis for tackling research waste. *Nature Ecology & Evolution*, *4*(4), 495–497. https://doi.org/10.1038/s41559-020-1141-6

Grekov, P., Pustejovsky, J. E., & Klingbeil, D. A. (2025). Flexible distributional models for meta-analysis of reading fluency outcomes from single-case designs: An examination using Bayesian methods. *Journal of School Psychology*, *110*, 101429. https://doi.org/10.1016/j.jsp.2025.101429

Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., et al. (2008). GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*, *336*, 924–92. https://doi.org/10.1136/bmj.39489.470347.AD

Heathers, J. A. J. (2025). *An introduction to forensic metascience* [Preprint]. https://doi.org/10.5281/zenodo.14871843

Henry, L., & Winfield, J. (2010). Working memory and educational achievement in children with intellectual disabilities. *Journal of Intellectual Disability Research*, *54*(4), 354–365. https://doi.org/10.1111/j.1365-2788.2010.01264.x

Hepburn, L., Beamish, W., & Alston-Knox, C. L. (2021). High school teacher perceptions of implementation of evidence-based practices for classroom management. *Teacher Development*, *25*(5), 669–686. https://doi.org/10.1080/13664530.2021.1957007

Horner, R. H., Carr, E. G., Halle, J. W., McGee, G. G., Odom, S. L., & Wolery, M. (2005). The Use of Single-Subject Research to Identify Evidence-Based Practice in Special Education: *Exceptional Children*, *71*(2), 165–179. https://doi.org/10.1177/001440290507100203

Hornung, R., Németh, L., Zadorozhnyi, O., Ullmann, T., Kammer, M., Killick, R., Paciorek, C., Chiquet, J., Herrmann, M., Batinović, L., Carlsson, R., Neuvial, P., Hejblum, B., Wrobel, J., Boulesteix, A.-L., & Tabelow, K. (2026). Overcoming barriers to computational reproducibility. *Royal Society Open Science*, *13*(6), 252489. https://doi.org/10.1098/rsos.252489

Huang, C.-K., Neylon, C., Montgomery, L., Hosking, R., Diprose, J. P., Handcock, R. N., & Wilson, K. (2024). Open access research outputs receive more diverse citations. *Scientometrics*, *129*(2), 825–845. https://doi.org/10.1007/s11192-023-04894-0

Ioannidis, J. P. A., Fanelli, D., Dunne, D. D., & Goodman, S. N. (2015). Meta-research: Evaluation and Improvement of Research Methods and Practices. *PLOS Biology*, *13(10)*. https://doi.org/doi.org/10.1371/journal.pbio.1002264

Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLOS Medicine*, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Kadlec, D., Sainani, K. L., & Nimphius, S. (2023). With Great Power Comes Great Responsibility: Common Errors in Meta-Analyses and Meta-Regressions in Strength & Conditioning Research. *Sports Medicine*, *53*(2), 313–325. https://doi.org/10.1007/s40279-022-01766-0

Khalil, H., Peters, M. D., McInerney, P. A., Godfrey, C. M., Alexander, L., Evans, C., Pieper, D., Moraes, E. B., Tricco, A. C., Munn, Z., & Pollock, D. (2022). The role of scoping reviews in reducing research waste. *Journal of Clinical Epidemiology*, *152*, 30–35. https://doi.org/10.1016/j.jclinepi.2022.09.012

Khodyakov, D., Grant, S., Kroger, J., & Bauman, M. (2023). *RAND Methodological Guidance for Conducting and Critically Appraising Delphi Panels*. RAND Corporation. https://doi.org/10.7249/TLA3082-1

Kok, S. I., Fetz, L. M., Stikkelbroek, Y. A. J., Leijten, P., Staaks, J., & Marić, M. (2026). A multilevel meta-analysis of single-case research on interventions for externalizing behavior problems in children and adolescents. *JAACAP Open*, *4*(2), 220–231. https://doi.org/10.1016/j.jaacop.2025.12.002

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D., & Shadish, W. R. (2012). Single-Case Intervention Research Design Standards. *Remedial and Special Education*, *34*(1), 26–38. https://doi.org/10.1177/0741932512452794

Lajeunesse, M. J. (2021, April). Juicr: Automated and Manual Extraction of Numerical Data from Scientific Images. Retrieved May 3, 2024, from https://cran.r-project.org/web/packages/juicr/index.html

Lakens, D. (2025a). Concerns About Replicability Across Two Crises in Social Psychology. *International Review of Social Psychology*, *38*(1). https://doi.org/10.5334/irsp.1036

Lakens, D. (2025b). Concerns About Theorizing, Relevance, Generalizability, and Methodology Across Two Crises in Social Psychology. *International Review of Social Psychology*, *38*(1). https://doi.org/10.5334/irsp.1038

Lakens, D. (2022). *Improving your statistical inferences*. https://lakens.github.io/statistical_inferences/

Lakens, D., Hilgard, J., & Staaks, J. (2016). On the reproducibility of meta-analyses: Six practical recommendations. *BMC Psychology*, *4*(24). https://doi.org/10.1186/s40359-016-0126-3

Lakens, D., Mesquida, C., Rasti, S., & Ditroilo, M. (2024, March). *The Benefits of Preregistration and Registered Reports* [Preprint]. https://doi.org/10.31234/osf.io/dqap7

Laverty, C., & Richards, C. (2026). Promoting accessible research for children with intellectual disabilities; lessons learnt from adaptations through the COVID-19 pandemic. *Journal of Intellectual Disability Research*, *70*(1), 68–75. https://doi.org/10.1111/jir.70055

Levin, H. M. (1988). Cost-Effectiveness and Educational Policy. *Educational Evaluation and Policy Analysis*, *10*(1), 51–69. https://doi.org/10.3102/01623737010001051

Lindsay, S., & Varahra, A. (2022). A systematic review of self-determination interventions for children and youth with disabilities. *Disability and Rehabilitation*, *44*(19), 5341–5362. https://doi.org/10.1080/09638288.2021.1928776

Liverpool, L. (2023). AI intensifies fight against 'paper mills' that churn out fake research. *Nature*, *618*(7964), 222–223. https://doi.org/10.1038/d41586-023-01780-w

Maggin, D. M., Barton, E., Reichow, B., Lane, K. L., & Shogren, K. A. (2022). Commentary on the What Works Clearinghouse Standards and Procedures Handbook (v. 4.1) for the Review of Single-Case Research.

Remedial and Special Education, *43*(6), 421–433. https://doi.org/10.1177/07419325211051317

Makel, M. C., Cook, B., Mu, N., Stegenga, S. M., Wong, V. C., & Drummond, K. (2024, December). *Beyond individual studies: A case for team science through systematic replications* [Preprint]. https://doi.org/10.35542/osf.io/ndzjc

Makel, M. C., Hodges, J., Cook, B. G., & Plucker, J. A. (2021). Both Questionable and Open Research Practices Are Prevalent in Education Research. *Educational Researcher*, *50*(8), 493–504. https://doi.org/10.3102/0013189X211001356

Max Planck Society. (n.d.). Berlin Declaration. Retrieved February 25, 2026, from https://openaccess.mpg.de/Berlin-Declaration

McKenna, J. W., Solis, M., Garwood, J., & Parenti, M. (2024). Characteristics of Individualized Education Programs for Students With Learning Disabilities: A Systematic Review. *Learning Disability Quarterly*, *47*(3), 194–207. https://doi.org/10.1177/07319487231182697

McKenzie, K., Milton, M., Smith, G., & Ouellette-Kuntz, H. (2016). Systematic Review of the Prevalence and Incidence of Intellectual Disabilities: Current Trends and Issues. *Current Developmental Disorders Reports*, *3*(2), 104–115. https://doi.org/10.1007/s40474-016-0085-7

Meehl, P. E. (1967). Theory-Testing in Psychology and Physics: A Methodological Paradox. *Philosophy of Science*, *34*(2), 103–115. https://doi.org/10.1086/288135

Meehl, P. E. (1990). Why Summaries of Research on Psychological Theories are Often Uninterpretable. *Psychological Reports*, *66*(1), 195–244. https://doi.org/10.2466/pr0.1990.66.1.195

Meta-psychology. (2023). Preregistration checklist [Document]. https://osf.io/t54ph/files/jt9v7

Miske, O., Abatayo, A. L., Daley, M., Dirzo, M., Fox, N., Haber, N., Hahn, K. M., Struhl, M. K., Mawhinney, B., Silverstein, P., Stankov, T., Tyner, A. H., Adamkovič, M., Alzahawi, S., Anafinova, S., Awtrey, E., Axxe, E., Bailey, J., Bakker, B. N., … Errington, T. M. (2026). Investigating the reproducibility of the social and behavioural sciences. *Nature*, *652*(8108), 126–134. https://doi.org/10.1038/s41586-026-10203-5

Mol, B. W., Lai, S., Rahim, A., et al. (2023). Checklist to assess trustworthiness in RAndomised Controlled Trials (TRACT checklist): Concept proposal and pilot. *Research Integrity and Peer Review*, *8*, 6. https://doi.org/10.1186/s41073-023-00130-8

Moretti, M., Alves, I., & Maxwell, G. (2012). A Systematic Literature Review of the Situation of the International Classification of Functioning, Disability, and Health and the International Classification of Functioning, Disability, and Health–Children and Youth Version in Education: A Useful Tool or a Flight of Fancy? *American Journal of Physi-*

cal Medicine & Rehabilitation, *91*(13), S103–S117. https://doi.org/10.1097/PHM.0b013e31823d53b2

Murad, M. H., Asi, N., Alsawas, M., & Alahdab, F. (2016). New evidence pyramid. *Evidence-Based Medicine*, *21*(4), 125–127. https://doi.org/10.1136/ebmed-2016-110401

Nordström, T., Kalmendal, A., & Batinovic, L. (2023). Risk of bias and open science practices in systematic reviews of educational effectiveness: A meta-review. *Review of Education*, *11*(3), e3443. https://doi.org/10.1002/rev3.3443

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., … Yarkoni, T. (2015). Promoting an open research culture. *Science*, *348*(6242), 1422–1425. https://doi.org/10.1126/science.aab2374

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability. *Perspectives on Psychological Science*, *7*(6), 615–631. https://doi.org/10.1177/1745691612459058

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. https://doi.org/10.1177/001440290507100201

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

Patall, E. A. (2021). Implications of the open science era for educational psychology research syntheses. *Educational Psychologist*, *56*(2), 142–160. https://doi.org/10.1080/00461520.2021.1897009

Patel, D. R., Cabral, M. D., Ho, A., & Merrick, J. (2020). A clinical primer on intellectual disability. *Translational Pediatrics*, *9*(Suppl 1), 23–35. https://doi.org/10.21037/tp.2020.02.02

Pearson, K. (1904). Report on Certain Enteric Fever Inoculation Statistics. *British Medical Journal*, *2*(2288), 1243–1246.

Pellegrini, M., Pigott, T. D., Chubb, C. S., Pruitt, N., & Scarbrough, H. F. (2025). Exploring the Methodological Quality of Education Interven-

tion Meta-Analyses: A Meta-Review. *Review of Educational Research*. https://doi.org/10.3102/00346543251389938

Perry, T., Morris, R., & Lea, R. (2022). A decade of replication study in education? A mapping review (2011–2020). *Educational Research and Evaluation*, *27*(1-2), 12–34. https://doi.org/10.1080/13803611.2021.2022315

Peters, M. D. J., Marnie, C., Tricco, A. C., Pollock, D., Munn, Z., Alexander, L., McInerney, P., Godfrey, C. M., & Khalil, H. (2020). Updated methodological guidance for the conduct of scoping reviews. *JBI Evidence Synthesis*, *18*(10), 2119–2126. https://doi.org/10.11124/JBIES-20-00167

Pustejovsky, J. E., & Ferron, J. M. (2017). Research Synthesis and Meta-Analysis of Single-Case Designs. In J. M. Kauffman, D. P. Hallahan, & P. C. Pullen (Eds.), *Handbook of Special Education* (2nd ed., pp. 168–186). Routledge. https://doi.org/10.4324/9781315517698-15

R Core Team. (2024). *R: A language and environment for statistical computing* (manual). R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Riley, R. D., Ensor, J., Hattle, M., Papadimitropoulou, K., & Morris, T. P. (2023). Two-stage or not two-stage? That is the question for IPD meta-analysis projects. *Research Synthesis Methods*, *14*(6), 903–910. https://doi.org/10.1002/jrsm.1661

Ropovik, I., Adamkovic, M., & Greger, D. (2021). Neglect of publication bias compromises meta-analyses of educational research. *PLoS ONE*, *16*(6), e0252415. https://doi.org/10.1371/journal.pone.0252415

Rosengaard, L. O., Andersen, M. Z., Rosenberg, J., & Fonnes, S. (2024). Five aspects of research waste in biomedicine: A scoping review. *Journal of Evidence-Based Medicine*, *17*(2), 351–359. https://doi.org/10.1111/jebm.12616

Sanderson, K. (2024). Science's fake-paper problem: High-profile effort will tackle paper mills. Retrieved May 7, 2026, from https://www.nature.com/articles/d41586-024-00159-9

Scheel, A. M. (2022). Why most psychological research findings are not even wrong. *Infant and Child Development*, *31*(1), e2295. https://doi.org/10.1002/icd.2295

Schor, S., & Karten, I. (1966). Statistical evaluation of medical journal manuscripts. *JAMA*, *195*(13), 1123–1128. https://doi.org/10.1001/jama.1966.03100130097026

Shogren, K. A., Palmer, S. B., Wehmeyer, M. L., Williams-Diehm, K., & Little, T. D. (2012). Effect of Intervention With the Self-Determined Learning Model of Instruction on Access and Goal Attainment. *Remedial and Special Education*, *33*(5), 320–330. https://doi.org/10.1177/0741932511410072

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Simonsohn, U., Simmons, J., & Nelson, L. (2023, June). *[109] Data Falsificada (Part 1): "Clusterfake"*. Data Colada. https://datacolada.org/109

Sims, S., Anders, J., Inglis, M., & Lortie-Forgues, H. (2023). Quantifying "Promising Trials Bias" in Randomized Controlled Trials in Education. *Journal of Research on Educational Effectiveness*, *16*(4), 663–680. https://doi.org/10.1080/19345747.2022.2090470

Slocum, T. A., Joslyn, P. R., Nichols, B., & Pinkelman, S. E. (2022). Revisiting an Analysis of Threats to Internal Validity in Multiple Baseline Designs. *Perspectives on Behavior Science*, *45*(3), 681–694. https://doi.org/10.1007/s40614-022-00351-0

Southward, J. D., & Kyzar, K. (2017). Predictors of competitive employment for students with intellectual and/or developmental disabilities. *Education and Training in Autism and Developmental Disabilities*, *52*(1), 26–37. https://doi.org/10.1177/215416471705200104

Spaniol, M., & Danielsson, H. (2022). A meta-analysis of the executive function components inhibition, shifting, and attention in intellectual disabilities. *Journal of Intellectual Disability Research*, *66*(1-2), 9–31. https://doi.org/10.1111/jir.12878

Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H.-Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., … Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *BMJ*, *366*. https://doi.org/10.1136/bmj.l4898

Stewart, L. A., Clarke, M., Rovers, M., Riley, R. D., Simmonds, M., Stewart, G., & Tierney, J. F. (2015). Preferred Reporting Items for a Systematic Review and Meta-analysis of Individual Participant Data: The PRISMA-IPD Statement. *JAMA*, *313*(16), 1657–1665. https://doi.org/10.1001/jama.2015.3656

Struening, E. L., & Guttentag, M. (Eds.). (1975). *Handbook of evaluation research* (Third edition). SAGE.

The Methods Group of the Campbell Collaboration. (2016, December). *Methodological expectations of Campbell Collaboration intervention reviews: Conduct standards* (tech. rep.). The Campbell Collaboration. https://doi.org/10.4073/cpg.2016.3

Thomas, J., Noel-Storr, A., Marshall, I., Wallace, B., McDonald, S., Mavergames, C., Glasziou, P., Shemilt, I., Synnot, A., Turner, T., Elliott, J., & Living Systematic Review Network. (2017). Living systematic reviews: 2. Combining human and machine effort. *Journal of Clin-*

ical Epidemiology, *91*, 31–37. https://doi.org/10.1016/j.jclinepi.2017.08.011

Tian, H., & Liu, Y. (2025). Effect of executive function on mathematical ability in children with intellectual disabilities. *International Journal of Developmental Disabilities*. https://doi.org/10.1080/20473869.2025.2520272

Tilburg University. (2012). *Flawed Science: The Fraudulent Research Practices of Social Psychologist Diederik Stapel* (tech. rep.). Tilburg University.

Topor, M., & Danielsson, H. (2026). *A Call to Action: Open Assessment Measures in Developmental Psychology* [Preprint]. https://doi.org/10.31234/osf.io/bazys_v1

Tsuji, S., Bergmann, C., & Cristia, A. (2014). Community-Augmented Meta-Analyses: Toward Cumulative Data Assessment. *Perspectives on Psychological Science*, *9*(6), 661–665. https://doi.org/10.1177/1745691614552498

Tungate, A. S., & Conners, F. A. (2021). Executive function in Down syndrome: A meta-analysis. *Research in Developmental Disabilities*, *108*, 103802. https://doi.org/10.1016/j.ridd.2020.103802

Ulriksen, L. B., Bilet-Mossige, M., Cogo-Moreira, H., Øien, R., & Nordahl-Hansen, A. (2024). Reading instruction for students with intellectual disabilities who require augmentative and alternative communication: A multiple single case study with baseline, posttest, follow-up, and maintenance. *Research in Developmental Disabilities*, *151*, 104790. https://doi.org/10.1016/j.ridd.2024.104790

UNESCO. (2006). Convention on the Rights of Persons with Disabilities. Retrieved December 24, 2025, from https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-persons-disabilities

UNESCO. (2021, December). UNESCO recommendation on open science. https://doi.org/10.54677/MNMH8546

van Dijk, W., Schatschneider, C., & Hart, S. A. (2021). Open Science in Education Sciences. *Journal of Learning Disabilities*, *54*(2), 139–152. https://doi.org/10.1177/0022219420945267

van den Akker, O. R., Peters, G.-J. Y., Bakker, C. J., Carlsson, R., Coles, N. A., Corker, K. S., Feldman, G., Moreau, D., Nordström, T., Pickering, J. S., Riegelman, A., Topor, M. K., van Veggel, N., Yeung, S. K., Call, M., Mellor, D. T., & Pfeiffer, N. (2023). Increasing the transparency of systematic reviews: Presenting a generalized registration form. *Systematic Reviews*, *12*(1), 170. https://doi.org/10.1186/s13643-023-02281-7

van der Molen, M., van Luit, J., van der Molen, M., Klugkist, I., & Jongmans, M. (2010). Effectiveness of a computerised working memory training in adolescents with mild to borderline intellectual disabilities. *Jour-*

nal of Intellectual Disability Research, *54*(5), 433–447. https://doi.org/10.1111/j.1365-2788.2010.01285.x

Vicente-Saez, R., & Martinez-Fuentes, C. (2018). Open Science now: A systematic literature review for an integrated definition. *Journal of Business Research*, *88*, 428–436. https://doi.org/10.1016/j.jbusres.2017.12.043

Wald, N., Harland, T., & Daskon, C. (2024). The gap statement and justification in higher education research: An analysis of published articles. *European Journal of Higher Education*, *14*(2), 308–323. https://doi.org/10.1080/21568235.2023.2189132

Wilkinson, J., Bero, L., Heal, C., Flemyng, E., Antoniou, G. A., Aburrow, T., Alfirevic, Z., Avenell, A., Barbour, V., Berghella, V., Bishop, D. V. M., Bordewijk, E. M., Brown, N. J. L., Christopher, J., Clarke, M., Dahly, D., Dennis, J., Dicker, P., Dumville, J., … Kirkham, J. J. (2026). INSPECT-SR tool for assessing trustworthiness of randomised controlled trials. *BMJ*, *394*, e100611. https://doi.org/10.1136/bmj-2026-100611

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18

Williamson, R. L., Jasper, A., Novak, J., Smith, C., Hunter, W., Casey, L., & Reeves, K. (2019). Re-examining Evidence Based Practice in Special Education: A Discussion. *Journal of International Special Needs Education*, *22*(2), 54–65. https://doi.org/10.9782/17-00022

Wolf, B. (2025). What Works for Whom: Exploring the Students, Settings, and Outcomes in What Works Clearinghouse Study Data. *Journal of Research on Educational Effectiveness*. https://doi.org/10.1080/19345747.2024.2427762

Wong, M. H. I. (2025). A Bayesian approach to small samples: Mixed-effects modeling in L2 interventional research. *Research Methods in Applied Linguistics*, *4*(3), 100231. https://doi.org/10.1016/j.rmal.2025.100231

World Health Organization. (n.d.). ICD-11 for Mortality and Morbidity Statistics. Retrieved December 24, 2025, from https://icd.who.int/browse/2025-01/mms/en#605267007

Zablotsky, B., Ng, A. E., Black, L. I., & Blumberg, S. J. (2023). *Diagnosed developmental disabilities in children aged 3–17 years: United States, 2019–2021* (NCHS Data Brief No. 473). National Center for Health Statistics. https://doi.org/10.15620/cdc:129520

Acknowledgments

I have been so lucky to go through this journey supported by incredible mentors, who made the work nothing but enjoyable (which left the misery for me to create all on my own).

To Rickard, for opening so many doors for me. Thank you for believing in me so much that you made so many others follow suit. To Henrik, for letting me walk through those opened doors on my own, safely and confidently, because I knew you were always right next to me. To Marta, for co-leading the way and helping me push through the most stressful moments. To Chelsea, for being available beyond your requirements and offering support out of sheer joy of helping people to grow and learn. I further thank Samuel Odom and Melby-Lervåg for their invaluable input that helped shape this dissertation.

To the colleagues at the division and friends outside. To Cristina, for the wonders of DnD, collaborative efforts to integrate witchcraft in academia and endless shitposting about work. To Francesca, for sharing the love for sun, warmth and great food that kept the sense of familiarity away from home. To Emil, who was always ready to avoid answering difficult and annoying questions I had, but made the struggles easier nonetheless. To Theres, Karin and Josefine, for the selfless help offered during challenging moments. And to Rodrigo, for the support and fun times throughout the years.

To colleagues at ELD, for creating the first sense of belonging outside of home. To André, for paving the way for weird PhD students that made it easy for me to follow. To Thomas, who offered wisdom and encouragement along the way. And to Natalie, for the heavily needed sanity checks.

To Marlon, for keeping me entertained with endless discussions about Beyonce, world orientation, and the importance of verbs and objects in sentences. (And to Beyonce, for Renaissance helped me go through data extraction without killing myself)

Anamariji i Danici, za neprekidno shitpostanje i najugodnija druženja, i prijateljstvo koje je tu neovisno o mojim poslovnim postignućima jer ni u ovom trenutku nećete znati o čemu se radi,

a svejedno ćete mi biti podrška. Niki, za prijateljstvo zbog kojeg sam uspjela dogurati do doktorata.

Mojoj obitelji, teti, sestri i mojim roditeljima, koji su mi omogućili život u kojem nijedan pothvat nije nedostižan ili zastrašujuć, pa tako ni ovaj ni u jednom trenutku nije bio. Hvala vam što ste uvijek u zadnjem vagonu vlaka.

Juri. Hvala ti za to što sam ja danas tu.

Sebi. Zahvaljujem sebi što sam ustrajna, što neprestano tražim više, što uživam u izazovima i želim naučiti novo i osvojiti teže. Zahvaljujem na tome što sam zbog toga sebe dovela do i provela kroz ovo iskustvo.

Nikad više, ali bilo je lijepo.

Studies from the Swedish Institute for Disability Research

1. **Varieties of reading disability**
   Stefan Gustafson
   ISBN 91-7219-867-2, 2000

2. **Cognitive functions in drivers with brain injury – anticipation and adaptation**
   Anna Lundqvist
   ISBN 91-7219-967-9, 2001

3. **Cognitive deafness**
   Ulf Andersson
   ISBN 91-7373-029-7, 2001

4. **Att lära sig leva med förvärvad hörselnedsättning sett ur par-perspektiv**
   Carin Fredriksson
   ISBN 91-7373-105-6, 2001

5. **Signs, Symptoms, and Disability Related to the Musculo-Skeletal System**
   Gunnar Lundberg
   ISBN 91-7373-160-9, 2002

6. **Participation – Ideology and Everyday Life**
   Anette Kjellberg
   ISBN 91-7373-371-7, 2002

7. **Föräldrar med funktionshinder – om barn, föräldraskap och familjeliv**
   Marie Gustavsson Holmström
   ISBN 91-7203-500-5, 2002

8. **Active wheelchair use in daily life**
   Kersti Samuelsson
   ISBN 91-7373-196-X, 2002

9. **Två kön eller inget alls. Politiska intentioner och vardagslivets realiteter i den arbetslivsinriktade rehabiliteringen**
   Marie Jansson
   ISBN 91-7373-568-X, 2003

10. **Audiological and cognitive long-term sequelae from closed head injury**

    Per-Olof Bergemalm

    ISBN 91-7668-384-2, 2004

11. **Att vara i särklass – om delaktighet och utanförskap i gymnasiesärskolan**

    Martin Molin

    ISBN 91-85295-46-9, 2004

12. **Rättvis idrottsundervisning för elever med rörelsehinder – dilemma kring omfördelning och erkännande**

    Kajsa Jerlinder

    Licentiate Degree, 2005

13. **Hearing impairment and deafness. Genetic and environmental factors – interactions – consequences. A clinical audiological approach**

    Per-Inge Carlsson

    ISBN 91-7668-426-1, 2005

14. **Hearing and cognition in speech comprehension. Methods and applications**

    Mathias Hällgren

    ISBN 91-85297-93-3, 2005

15. **Living with deteriorating and hereditary disease: experiences over ten years of persons with muscular dystrophy and their next of kin**

    Katrin Boström

    ISBN 91-7668-427-x, 2005

16. **Disease and disability in early rheumatoid arthritis**

    Ingrid Thyberg

    ISBN 91-85299-16-2, 2005

17. **"Varför får jag icke följa med dit fram?" Medborgarskapet och den offentliga debatten om dövstumma och blinda 1860-1914**

    Staffan Bengtsson

    ISBN 91-85457-06-X, 2005

18. **Modalities of Mind. Modality-specific and nonmodality-specific aspects of working memory for sign and speech**

    Mary Rudner

    ISBN 91-85457-10-8, 2005

19. **Facing the Illusion Piece by Piece. Face recognition for persons with learning disability**

Henrik Danielsson

ISBN 91-85497-09-6, 2006

20. **Vuxna med förvärvad traumatisk hjärnskada – omställningsprocesser och konsekvenser i vardagslivet. En studie av femton personers upplevelser och erfarenheter av att leva med förvärvad traumatisk hjärnskada**

Thomas Strandberg

ISBN 91-7668-498-9, 2006

21. **Nycklar till kommunikation. Kommunikation mellan vuxna personer med grav förvärvad hjärnskada och personernas närstående, anhöriga och personal**

Pia Käcker

ISBN 978-91-85715-88-6, 2007

22. **”Aspergern, det är jag”. En intervjustudie om att leva med Asperger syndrom**

Gunvor Larsson Abbad

ISBN 978-91-85831-43-2, 2007

23. **Sounds of silence – Phonological awareness and written language in children with and without speech**

Janna Ferreira

ISBN 978-91-85895-74-8, 2007

24. **Postponed Plans: Prospective Memory and Intellectual Disability**

Anna Levén

ISBN 978-91-85895-57-1, 2007

25. **Consequences of brain tumours from the perspective of the patients and of their next of kin**

Tanja Edvardsson

ISBN 978-91-7668-572-3, 2008

26. **Impact on participation and service for persons with deafblindness**

Kerstin Möller

ISBN 978-91-7668-595-2, 2008

27. **Approaches to Audiological Rehabilitation with Hearing Aids: studies on prefitting strategies and assessment of outcomes**

Marie Öberg

ISBN 978-91-7393-828-0, 2008

28. **Social Interaction and Participation in Activities of Everyday Life Among Persons with Schizophrenia**

    Maria Yilmaz

    Licentiate Degree, 2009

29. **Focus on Chronic Disease through Different Lenses of Expertise Towards Implementation of Patient-Focused Decision Support Preventing Disability:**

    The example of Early Rheumatoid Arthritis

    Örjan Dahlström

    ISBN 978-91-7393-613-2, 2009

30. **Children with Cochlear Implants: Cognition and Reading Ability**

    Malin Wass

    ISBN: 978-91-7393-487-9, 2009

31. **Restricted participation: Unaccompanied children in interpreter-mediated asylum hearings in Sweden**

    Olga Keselman

    ISBN: 978-91-7393-499-2, 2009

32. **Deaf people and labour market in Sweden. Education – Employment – Economy**

    Emelie Rydberg

    ISBN: 978-91-7668-725-3, 2010

33. **Social rättvisa i inkluderande idrottsundervisning för elever med rörelsehinder – en utopi?**

    Kajsa Jerlinder

    ISBN: 978-91-7668-726-0, 2010

34. **Erfarenheter av rehabiliteringsprocessen mot ett arbetsliv – brukarens och de professionellas perspektiv**

    Helene Hillborg

    ISBN: 978-91-7668-741-3, 2010

35. **Knowing me, knowing you – Mentalization abilities of children who use augmentative and alternative communication**

    Annette Sundqvist

    ISBN: 978-91-7393-316-2, 2010

36. **Lärare, socialsekreterare och barn som far illa – om sociala representationer och interprofessionell samverkan**

Per Germundsson

ISBN: 978-91-7668-787-1, 2011

37. **Fats in Mind**
Effects of Omega-3 Fatty Acids on Cognition and Behaviour in Childhood

Ulrika Birberg Thornberg

ISBN: 978-91-7393-164-9, 2011

38. **”Jobbet är kommunikation”**
Om användning av arbetshjälpmedel för personer med hörselnedsättning

Sif Bjarnason

Licentiate Degree. ISBN: 978-91-7668-835-9, 2011

39. **Applying the ICF-CY to identify everyday life situations of children and youth with disabilities**

Margareta Adolfsson

ISBN: 978-91-628-8342-3, 2011

40. **Tinnitus – an acceptance-based approach**

Vendela Zetterqvist

ISBN: 978-91-7393-040-6, 2011

41. **Applicability of the ICF-CY to describe functioning and environment of children with disabilities**

Nina Klang

ISBN: 978-91-7668-864-9, 2012

42. **Bringing more to participation**
Participation in school activities of persons with Disability within the framework of the International Classification of Functioning, Disability and Health for Children and Youth (ICF-CY)

Gregor Maxwell

ISBN: 978-91-628-8484-0, 2012

43. **From Eye to Us.**
Prerequisites for and levels of participation in mainstream school of persons with Autism Spectrum Conditions

Marita Falkmer

ISBN: 978-91-637-2091-8, 2013

44. **Otosclerosis, clinical long-term perspectives**

Ylva Dahlin-Redfors

ISBN 978-91-628-8617-2, 2013

45. **Tinnitus in Context - A Contemporary Contextual Behavioral Approach**

Hugo Hesser

ISBN 978-91-7519-701-2, 2013

46. **Hearing and middle ear status in children and young adults with cleft palate**

Traci Flynn

ISBN 978-91-628-8645-5, 2013

47. **Utrymme för deltagande, beslutsprocesser i möten mellan patienter med ospecifika ländryggsbesvär och sjukgymnaster i primär vård**

Iréne Josephson

ISBN 42-978-91-85835-41-6, 2013

48. **”Man vill ju klara sig själv” Studievardagen för studenter med Asperger syndrom i högre studier**

Ann Simmeborn Fleischer

ISBN 978-91-628-8681-3, 2013

49. **Cognitive erosion and its implications in Alzheimer’s disease**

Selina Mårdh

ISBN 978-91-7519-621-1, 2013

50. **Hörselscreening av en population med utvecklingsstörning**
Utvärdering av psykoakustisk testmetod och av OAE-registrering som komplementär metod

Eva Andersson

Licentiate Degree. ISBN 978-91-7519-616-9, 2013

51. **Skolformens komplexitet – elevers erfarenheter av skolvardag och tillhörighet i gymnasiesärskolan**

Therése Mineur

ISBN 978-91-7668-951-6, 2013

52. **Evaluating the process of change:**
Studies on patient journey, hearing disability acceptance and stages-of-change

Vinaya Kumar Channapatna Manchaiah

ISBN 978-91-7519-534-6, 2013

53. **Cognition in hearing aid users: Memory for everyday speech**

Hoi Ning (Elaine) Ng

ISBN 978-91-7519-494-3, 2013

54. **Representing sounds and spellings Phonological decline and compensatory working memory in acquired hearing impairment**

Elisabet Classon

ISBN 978-91-7519-500-1, 2013

55. **Assessment of participation in people with a mild intellectual disability**

Patrik Arvidsson

ISBN 978-91-7668-974-5, 2013

56. **Barnperspektiv i barnavårdsutredningar – med barns hälsa och barns upplevelser i fokus**

Elin Hultman

ISBN 978-91-7519-457-8, 2013

57. **Internet Interventions for Hearing Loss**
Examining rehabilitation Self-report measures and Internet use in hearing-aid users

Elisabet Sundewall Thorén

ISBN 978-91-7519-423-3, 2014

58. **Exploring Cognitive Spare Capacity: Executive Processing of Degraded Speech**

Sushmit Mishra

ISBN 978-91-7519-386-1, 2014

59. **Supported employment i en svensk kontext – förutsättningar när personer med funktionsnedsättning når, får och behåller ett arbete**

Johanna Gustafsson

ISBN 978-91-7529-012-6, 2014

60. **Effects of Specific Cochlear Pathologies on the Auditory Functions: Modelling, Simulations and Clinical Implications**

Amin Saremi

ISBN 978-91-7519-365-6, 2014

61. **Children with profound intellectual and multiple disabilities and their participation in family activities**

Anna Karin Axelsson

ISBN 978-91-85835-48-5, 2014

62. **Lexical and Semantic Development in Children With Cochlear Implants**

Ulrika Löfkvist

ISBN 978-91-7549-546-0, 2014

63. **Rethinking sound. Computer-assisted reading intervention with a phonics approach for deaf and hard of hearing children using cochlear implants or hearing aids**

Cecilia Nakeva von Mentzer

ISBN 978-91-7519-270-3, 2014

64. **Assessing cognitive spare capacity as a measure of listening effort using the Auditory Inference Span Test**

Niklas Rönnberg

ISBN 978-91-7519-267-3, 2014

65. **Employees with Aided Hearing Impairment: An Interdisciplinary Perspective**

Håkan Hua

ISBN 978-91-7519-240-6, 2014

66. **Prosthetic and Orthotic Services in Developing Countries**

Lina Magnusson

ISBN 978-91-85835-55-3, 2014

67. **Dealing with digits - Arithmetic, memory and phonology in deaf signers**

Josefine Andin

ISBN: 978-91-7519-235-2, 2014

68. **Time is of the essence in speech recognition: Get it fast or think about it**

Shahram Moradi

ISBN: 978-91-7519-188-1, 2014

69. **Effects of hearing loss on traffic safety and mobility**

Birgitta Thorslund

ISBN: 978-91-7519-178-2, 2014

70. **Aspekter på lärande vid dövblindhet - möjligheter och begränsningar för personer med Alström syndrom**

Berit Rönnåsen

Licentiate Degree. ISBN: 978-91-7529-068-3, 2015

71. **Memory and communication in typically developing infants and children with Autism Spectrum Disorder: Behavioral and electrophysiological indices**

Emelie Nordqvist

ISBN: 978-91-7519-078-5, 2015

72. **Cognitive capacities and composite cognitive skills in individuals with Usher syndrome type 1 and 2**

Cecilia Henricson

ISBN: 978-91-7685-999-5, 2015

73. **Functioning and Disability in Adults with Hearing Loss Preparatory studies in the ICF Core Sets for Hearing Loss project**

Sarah Granberg

ISBN: 978-91-7529-086-7, 2015

74. **Speech recognition and memory processes in native and non-native language perception**

Lisa Kilman

ISBN: 978-91-7685-972-8, 2015

75. **To include or not to include: Teachers' social representations of inclusion of students with Asperger diagnosis**

Ann-Charlotte Linton

ISBN: 978-91-7685-956-8, 2015

76. **Health and People with Usher syndrome**

Moa Wahlqvist

ISBN: 978-91-7529-098-0, 2015

77. **Children with mild intellectual disability and their families – needs for support, service utilisation and experiences of support**

Lena Olsson

ISBN: 978-91-85835-66-9, 2016

78. **Deafblindness**
Theory-of-mind, cognitive functioning and social network in Alström syndrome

Hans-Erik Frölander

ISBN: 978-91-7529-132-1, 2016

79. **Signs for Developing Reading**
Sign Language and Reading Development in Deaf and Hard-of-Hearing Children

Emil Holmer

ISBN: 978-91-7685-767-0, 2016

80. **Speech masking speech in everyday communication**
The role of inhibitory control and working memory capacity

Victoria Stenbäck

ISBN: 978-7685-604-8, 2017

81. **Neural and Cognitive Effects of Hearing Loss on Speech Processing**

Eline Borch Petersen

ISBN: 978-91-7685-640-6, 2017

82. **A biopsychosocial approach to functioning, oral health and specialist dental health care in children with disabilities – Swedish and international perspectives**

Johanna Norderyd

ISBN: 978-91-85835-81-2, 2017

83. **Samverkansprojekt, och sen då? – en uppföljande studie av samverkansprocessen kring barn och unga som far illa eller riskerar att fara illa**

Ulrika Englund

ISBN: 978-91-7529-194-9, 2017

84. **Aural rehabilitation programs for hearing aid users**
Evaluating and clinically applying educational programs, supported via telephone and/or the internet and professionally guided by an audiologist

Milijana Malmberg

ISBN: 978-91-629-0187-5, 2017

85. **Tolkning vid förmedlade samtal via Bildtelefoni.net**
– interaktion och gemensamt meningsskapande

Camilla Warnicke

ISBN: 978-91-7529-197-0, 2017

86. **Vardagslivets aktiviteter. Handlingen som terapeutiskt redskap för personer med svårare psykiska funktionshinder – Analys av arbetsterapeuters berättelser med utgångspunkt i G H Meads teori om social handling**

Marianne Boström

ISBN: 978-91-7529-206-9, 2017

87. **Creative disability classification systems: The case of Greece, 1990-2015**

Antonia Pavli

ISBN: 978-91-7529-204-5, 2017

88. Ej utgiven

89. **Perception of disturbing sounds**
Investigations of people with hearing loss and normal hearing

Åsa Skagerstrand

ISBN: 978-91-7529-229-8, 2018

90. **Peer Interaction in Preschool: Necessary, but not Sufficient**
The Influence of Social Interaction on the Link between Behaviour Difficulties and Engagement among Children with and without Need of Special Support

Madeleine Sjöman

ISBN: 978-91-88339-13-3, 2018

91. **Prosodic and Phonological Ability in Children with Developmental Language Disorder and Children with Hearing Impairment**
In the context of Word and Nonword Repetition

Simon Sundström

ISBN: 978-91-7685-321-4, 2018

92. **From detection to intervention**
Psychological aspects of online hearing rehabilitation

Peter Molander

ISBN: 978-91-7685-226-2, 2018

93. **Participation in and outside school**
Self-ratings by Swedish adolescents with and without impairments and long-term health conditions

Frida Lygnegård

ISBN: 978-91-85835-92-8, 2018

94. **Återgång i arbete efter förvärvad hjärnskada**
– livskvalitet, möjligheter och hinder

Marie Matérne

ISBN: 978-91-7529-267-0, 2018

95. **Climbing up the hearing rehabilitation ladder**

Elisabeth Ingo

ISBN: 978-91-7519-011-2, 2019

96. **Situationen på den svenska arbetsmarknaden för personer med funktionsnedsättning**
En sekundäranalys om sysselsättning, yrkesnivå, diskriminering, stöd och anpassning för personer med olika typer av funktionsnedsättning

Tomas Boman

ISBN: 978-91-7529-302-8, 2019

97. **Different is cool!**
Self-efficacy and participation of students with and without disabilities in school-based Physical Education

Karin Bertills

ISBN: 978-91-88339-26-3, 2019

98. **Fonologisk utveckling hos barn med otitbenägenhet**
En longitudinell studie i åldrarna 3;6 till 5;6 år

Helena Stålnacke

Licentiate Degree. ISBN: 978-91-7529-321-9, 2020

99. **Life Strategies, Work and Health in People with Usher Syndrome**

Mattias Ehn

ISBN: 978-91-7529-322-6, 2020

100. **Time to Plan:**
How to support everyday planning in adolescents with intellectual disability

Lisa Palmqvist

ISBN: 978-91-7929-869-2, 2020

101. **Reasons for Language**
Language and Analogical Reasoning Ability in Children with Cochlear Implants and Children with Typical Hearing

Michaela Socher

ISBN: 978-91-7929-791-6, 2020

102. **The development of the Situated Phoneme (SiP) test:**
A Swedish test of phonemic discrimination in noise for adult people with hearing loss

Erik Witte

ISBN: 978-91-7529-369-1, 2021

103. **Cognition Seen Through the Eyes of Hearing Aid Users:**
Working Memory Resource Allocation for Speech Perception and Recall

Andreea Micula

ISBN: 978-91-7929-137-2, 2022

104. **Förändringar i transportrelaterad välfärd och myndighetsförtroende**
En studie om personer som fått sitt körkort återkallat på grund av synfältsbortfall

Jonna Nyberg

ISBN: 978-91-7529-413-1, 2021

105. **Adult patients with severe-to-profound hearing impairment**
A clinical, register-based, and interview study

Satu Turunen-Taheri

ISBN: 978-91-8016-338-5, 2021

106. **Older Adults with Dual Sensory Loss:**
Prevalence, Diagnoses & Rehabilitation Services

Elin Lundin

ISBN: 978-91-7529-437-7, 2022

107. **Auditory Distraction in ADHD: From Behaviour to the Brain**

Rina Blomberg

ISBN: 978-91-7929-302-4, 2022

108. **Assessment of everyday functioning for adults with hearing loss – Development of the Hearing and Functioning in Everyday Life Questionnaire (HFEQ)**

Elin Karlsson

ISBN: 978-91-7529-435-3, 2022

109. **Participation in physical activity among adolescents with autism spectrum disorder – the perspectives of adolescents, parents and professionals**

Susann Arnell

ISBN: 978-91-7529-461-2, 2022

Studies in Disability Research

110. **Words Don't Come Easy**
Decoding and Reading Comprehension Difficulties in Adolescents with Intellectual Disability

Karin Nilsson

ISBN: 978-91-7929-541-7, 2022

111. **An Exploration of Successful Aging: Integrating Perspectives and Exploring Disability Trends**

Andreea Badache

Licentiate Degree. ISBN:

112. **Participation in everyday activities among children with intellectual disabilities in mainland China: The development, verification, and application of the Picture My Participation instrument**

Lin Jun Shi

ISBN: 978-91-88669-33-9, 2023

113. **Everyday Life in Preschool - Swedish and International Approaches**

Frida Åström

ISBN: 978-91-88339-66-9, 2023

114. **Measurement of Child Engagement in Early Childhood Education and Care**

Andrea Ritosa

ISBN: 978-91-88339-69-0, 2023

115. **Finding a pathway – through a procedure in a high-technology environment for a child with an autism spectrum disorder from the perspectives of parents and healthcare professionals**

Emelie Pettersson

ISBN: 978-91-88669-45-2, 2024

116. **A gender perspective on Supported Employment**
Does gender influence access, process, and employment outcomes?

Ingrid Witte

ISBN: 978-91-7529-563-3, 2024

117. **Beyond the Years: How Sensory and Cognitive Functions in Old-Age Shape Disability, and Perspectives on Successful Aging: An epidemiological perspective of disability**

Andreea Badache

ISBN: 978-91-7529-635-7, 2025

118. **Vocal emotion identification in hearing loss**
Insights from emotion expression acoustics

Mattias Ekberg

ISBN: 978-91-8118-197-5, 2025

119. **Music and Hearing Health - A Study on Music Listening Behaviors and Hearing-Related Risks Among Young People**

Iris Elmazoska

ISBN: 978-91-7529-692-0, 2025

120. **Dialogues on Violence in Close Relationships:**
Perspectives from Professionals and Persons with Cognitive Disability

Sofie Adaszak

Licentiate Degree ISBN: 978-91-7529-724-8, 2025

Papers

The papers associated with this thesis have been removed for copyright reasons. For more details about these see:

https://doi.org/10.3384/9789181187670

FACULTY OF ARTS AND SCIENCES

Linköping Studies in Arts and Sciences No. 946, 2026

Department of Behavioural Sciences and Learning
Disability Research Division
Linköping University
SE-581 83 Linköping, Sweden

www.liu.se

LINKÖPING UNIVERSITY