Fidelity measurements in home-based intervention studies for children who stutter

Íris Ösp Bergþórsdóttir

Thesis for the degree of Philosophiae Doctor

March 2025

School of Health Sciences

FACULTY OF MEDICINE

UNIVERSITY OF ICELAND

Fidelity measurements in home-based intervention studies for children who stutter

Íris Ösp Bergþórsdóttir

Thesis for the degree of Philosophiae Doctor

Supervisor(s)

Jóhanna Thelma Einarsdóttir PhD
Kathryn Crowe PhD

Doctoral committee

Þorlákur Karlsson PhD

March 2025

School of Health Sciences

FACULTY OF MEDICINE

UNIVERSITY OF ICELAND

Mælingar á meðferðarheldni í heimaíhlutun fyrir börn sem stama

Íris Ösp Bergþórsdóttir

Ritgerð til doktorsgráðu

Leiðbeinandi/leiðbeinendur

Jóhanna Thelma Einarsdóttir PhD
Kathryn Crowe PhD

Doktorsnefnd

Þorlákur Karlsson PhD

Mars 2025

Heilbrigðisvísindasvið

LÆKNADEILD

HÁSKÓLI ÍSLANDS

ISBN (apply at http://www.landsbokasafn.is/)

ORCID: 0000-0002-5121-3503

Reykjavik, Iceland 2025

Ágrip

Bakgrunnur: Meðferðarheldni (e. fidelity of implementation) vísar til þess að hve miklu leyti íhlutun er gerð í samræmi við þá áætlun sem lagt var upp með. Mæling og eftirfylgni á meðferðarheldni er mikilvægur þáttur í íhlutunarrannsóknum til að greina og tryggja skilning á því hvað liggur að baki árangursríkrar íhlutunar. Til að meta árangur íhlutunar nota rannsakendur í auknum mæli mælingar á meðferðarheldni og öðlast þannig dýpri skilning á því hvernig íhlutunin var framkvæmd og hvort hún fylgdi áætlun.

Markmið: Aðalmarkmið þessarar ritgerðar var að skoða hvernig meðferðarheldni hefur áhrif á árangur í íhlutunarrannsóknum. Markmiðin voru að: 1) Fá yfirlit yfir hvernig greint er frá meðferðarheldni í íhlutunarrannsóknum á stami barna þar sem íhlutunin fer að stórum hluta til fram á heimili barnsins (grein I og II, rannsókn I). 2) Kanna árangur íhlutunarrannsóknar fyrir eldri börn sem stama, (stuttering intervention for older children [STOC]) (forrannsókn), þar sem nákvæmar mælingar á meðferðarheldni voru byggðar inn í rannsóknarsniðið og íhlutunin var gerð heima hjá barninu (grein III, rannsókn II). 3) Meta hvort hægt sé að ná góðri meðferðarheldni í íhlutunarrannsókn með því að gera ráð fyrir mælingum á meðferðarheldni í rannsóknarsniðinu og mæla þær ítarlega yfir íhlutunartímann (grein IV, rannsókn III). 4) Meta tengsl milli mælinga á meðferðarheldni og niðurstaðna á árangri íhlutunar í meðferð fyrir börn á skólaaldri sem stama (grein IV, rannsókn III).

Aðferð: Mismunandi aðferðir voru valdar eftir því um hvaða rannsóknarspurningu var að ræða. Til að fá yfirlit yfir meðferðarheldni í íhlutunarrannsóknum á stami barna var gerð kerfisbundin leit og fræðileg úttekt á ritrýndum greinum sem birtar voru á árunum 1981–2021. Þrjátíu og sex greinar uppfylltu skilyrði um nánari skoðun. Athugað var hvernig þær mældu meðferðarheldni. Meðferðarheldni var metin út frá fjórum lykilmælikvörðum, magni (e.dosage), heldni (e.adherence), gæðum (e.quality) og svörun (e.responsiveness) (grein II). Til að kanna árangur íhlutunar meðal barna sem stama var gerð einliða sniðsrannsókn með margliða grunnskeiði fyrir skólabörn (STOC). Náði rannsóknin til sjö barna á aldrinum 9 – 13 ára og var fylgst með stami þeirra í 18 til 30 mánuði, eftir því um hvaða barn var að ræða. Til að meta áhrif íhlutunar á talfærni barnanna var safnað mánaðarlegum málsýnum (n=653) og upptökum af hverri íhlutunarstund (n=1551) (grein III). Til að meta hvort hægt væri að ná góðri meðferðarheldni og hvort tengsl væru á milli mælinga á meðferðarheldni og niðurstaðna á árangri íhlutunar voru íhlutunarstundirnar (n=1551) metnar út frá fjórum þáttum meðferðarheldni: magni, heldni, gæðum og svörun. Meðferðarheldni var metin á þessum þáttum, bæði fyrir stundir sem foreldrar stýrðu og einnig fyrir stundir sem börnin stýrðu. Hver þáttur var sérstaklega skilgreindur og einnig þær aðferðir sem voru notaðar við matið (grein IV).

Niðurstöður: Í kerfisbundnu yfirlitsrannsókninni kom í ljós að engin rannsókn greindi frá öllum fjórum þáttum meðferðarheldni, hvorki í aðstæðum í klíník né heima. Fimm rannsóknir fjölluðu ekki um neinn þátt meðferðarheldni, í hvorugu umhverfinu. Fjöldi skráðra þátta í meðferðarheldni var frá 0 til 4 í klínískum aðstæðum (*M*=1.5) og frá 0 til 4 í aðstæðum heima. (*M*=1.0). Magn íhlutunnar var sá þáttur meðferðarheldni sem oftast var skráður (27 rannsóknir, 75,0%), á meðan svörun var sá þáttur sem var sjaldnast skráður (16 rannsóknir, 44,4%). Þrátt fyrir örlitla aukningu á því hvernig greint var frá meðferðarheldni í rannsóknum á síðustu árum sýndi heildarfjöldi skráðra þátta á meðferðarheldni ekki marktæka aukningu með tímanum (grein II). Forrannsóknin á STOC íhlutuninni sýndi jákvæðar niðurstöður fyrir alla þátttakendur, þeir stömuðu minna og alvarleiki stamsins minnkaði, ásamt því að þeir töluðu meira og á eðlilegri hátt. Þeir fjórir þátttakendur sem luku íhlutuninni sýndu marktækar breytingar, bæði hvað varðar talflæði og lífsgæði (grein III). Aukið magn íhlutunar tengdist því að þátttakendur stömuðu minna, töluðu meira og á eðlilegri hátt og stamið var ekki jafn alvarlegt. Átti það sérstaklega við um þær þjálfunarstundir sem börnin stýrðu. Meiri heldni tengdist minna stami, talið hljómaði eðlilegra og stam var ekki metið jafn alvarlegt. Hins vegar tengdust betri gæði og meiri svörun því að þátttakendur töluðu meira en hafði ekki áhrif á að stamið minnkaði, að talið hljómaði eðlilegra eða væri ekki eins alvarlegt (grein IV). Skráð stig fyrir alla fjóra þætti meðferðaheldni voru hærri meðal þeirra fjögra þátttakenda sem luku íhlutuninni en þeirra sem gerðu það ekki, óháð því hvort foreldrar eða börnin sáu um íhlutunarstundina (Grein IV).

Ályktanir: Ófullnægjandi skráning á meðferðarheldni í íhlutunarrannsóknum fyrir börn sem stama gerir það að verkum að krefjandi er fyrir meðferðaraðila og rannsakendur að túlka niðurstöður íhlutunarrannsókna á áreiðanlegan hátt. STOC forrannsóknin sýndi fram á meira talflæði og betri lífsgæði hjá þátttakendum sem luku íhlutuninni. Þátttakendur sem voru með betri meðferðarheldni í STOC rannsókninni komu betur út eða stömuðu minna. Mismunandi meðferðarheldni kom í ljós þegar hóparnir voru bornir saman eða þáttendur sem luku íhlutuninni voru bornir saman við þá sem gerðu það ekki. Þátttakendur sem luku íhlutuninni voru með betri meðferðarheldni en þeir sem kláruðu hana ekki. Ítarleg athugun á meðferðarheldni getur veitt nýja sýn á breytilega svörun við íhlutun í rannsóknum á stami í framtíðinni. Frekari rannsóknir eru nauðsynlegar til að kafa dýpra í ákveðna þætti STOC. Athugun á meðferðarheldni getur varpað ljósi á þá þætti sem kunna að leiða til mismunandi svörunar við íhlutun sem gerir það að forgangsatriði að mælingar og skráningu á meðferðarheldni séu nákvæmar.

Niðurstöður þessarar ritgerðar styðja við núverandi þekkingu um mikilvægi þess að mæla meðferðarheldni í íhlutunarrannsóknum. Að sniðganga mat á meðferðarheldni getur leitt til villandi tengsla milli íhlutunar og niðurstaðna, sem grefur undan áreiðanleika og notagildi rannsóknarniðurstaðna. Án gagna um meðferðarheldni gætu rannsakendur ályktað ranglega að íhlutun virkaði ekki, þegar raunveruleg virkni gæti verið töluverð ef íhlutunin hefur verið framkvæmd samkvæmt áætlun. Einnig gætu ítarleg gögn um meðferðaheldni sýnt að minna magn íhlutunar væri nægilegt. Auk þess veita ítarleg gögn um meðferðaheldni ómetanlegar upplýsingar um mælingar á niðurstöðum sem koma til vegna íhlutunarinnar og varpa ljósi á misræmi milli skynjaðra og raunverulegra áhrifa íhlutunarinnar.

Lykilorð:

Meðferðarheldni, íhlutun, börn, foreldrastýrð íhlutun, stam, talflæði.

Abstract

Background: Fidelity of Implementation (FOI) refers to the degree to which an intervention is delivered as intended or prescribed. Measuring and tracking FOI in intervention studies is an essential part of ensuring understanding of intervention effectiveness evaluations. Researchers are increasingly utilizing FOI to attain a more holistic comprehension of intervention delivery, aiming to assess whether implementation aligns with intended protocols and correlate this with intervention outcomes.

Aims: The overall aim of this thesis was to examine how implementation fidelity influences the effectiveness of an intervention. More specifically, the aims were to: 1) Gain a comprehensive understanding of the reporting of fidelity measurements in previous stuttering intervention research for children who stutter (CWS) where a component of the intervention was implemented in a home environment and to identify gaps in knowledge related to FOI measurements and reporting (Paper I & II, Study I). 2) Examine the effectiveness of a pilot home-based intervention for CWS (stuttering intervention for older children [STOC]) where detailed measurements of implementation fidelity used in the research design (Paper III, Study II). 3) Assess whether high levels of fidelity could be achieved in an intervention study where FOI was incorporated into the design and tightly controlled for during the study (Paper IV, Study III). 4) Determine the relationship between measured FOI and performance on outcome measures for an intervention for school-aged CWS (Paper IV, Study III).

Methods: The research was based on a range of methodologies where different methods were applied to answer different research questions. To address Aim 1, a systematic literature search and literature review of published, peer-reviewed papers published between 1981 until 2021 was completed. Thirty-six eligible papers were included and the presence of four key fidelity components in these papers were analysed: dosage, adherence, quality, and responsiveness (Paper II). To address Aim 2, a single case experimental design with multiple baselines was used. STOC intervention consisted of a home-based intervention program combining time-out training with a performance-contingent maintenance schedule strategy. This study was a pilot of the effect of the STOC intervention and involved seven male children aged 9 – 13 who stuttered with children's stuttering monitored for between 18 to 30 months, depending on the child. Routine monthly speech samples (*n*=653) and recordings of each intervention session (*n*=1551) were collected (Paper III) to determine the effect of the intervention on children's speech production. To address Aims 3 and 4, the intervention sessions (*n*=1551) from Paper III were evaluated based on the four FOI components: dosage, adherence, quality, and responsiveness. FOI was assessed across these dimensions for both parent- and child-managed STOC sessions, each component having specific definitions and methodologies employed for assessment (Paper IV).

Results: Within the literature review, no study reported on all four FOI components in both clinical and home settings, and five studies did not report on any FOI component in either setting. The reported number of FOI components ranged from 0 to 4 in clinical settings (*M*=1.5) and from 0 to 4 in home settings (*M*=1.0). Dosage was the most frequently reported FOI component (27 studies, 75.0%), while responsiveness was the least frequently reported (16 studies, 44.4%). Despite a slight trend towards increased reporting in recent years, the overall number of FOI components reported did not show a significant increase over time (Paper II). The STOC intervention pilot study demonstrated promising results for all participants, showing reductions in both the frequency and severity of stuttering, along with increases in speech quantity and naturalness. The four participants who completed the intervention reported significant improvements in both their fluency of speech and their quality of life (Paper III). Increased dosage was associated with reduced stuttering, increased speech output, enhanced naturalness, and reduced severity, particularly for sessions managed by the children. Greater adherence was linked to reduced stuttering, more natural speech, and lower severity ratings. However, higher quality and responsiveness were associated with increased speech output but did not correspond to reduced stuttering, enhanced naturalness, or decreased severity (Paper IV). Scores for all four FOI components were higher among participants who completed STOC compared to those who did not, regardless of whether sessions were managed by parents or children (Paper IV).

Conclusions: The poor reporting of FOI in intervention research for CWS poses a significant methodological challenge, hindering the ability of clinicians and researchers to interpret study findings accurately. The STOC pilot study revealed improvements in fluency and quality of life for the participants who completed the intervention. Participants who adhered more closely to STOC showed better outcomes; experiencing reduced stuttering. Discrepancies in FOI were apparent when comparing participants who completed the intervention with those who did not. A comprehensive examination of FOI may offer fresh insights into the variability in intervention response in future studies of stuttering interventions. Further research is warranted to delve deeper into the specifics of the STOC protocol. Exploring implementation fidelity could shed light on a significant factor contributing to the variability in intervention outcomes, making it imperative for intervention studies to prioritize thorough measurement and reporting of FOI to advance evidence-based practices.

The findings of this thesis contribute to the existing body of literature emphasizing the significance of FOI measurements in intervention research. Overlooking the assessment of FOI could lead to misleading associations between interventions and outcomes, undermining the reliability and applicability of study results. Without FOI data, researchers may inaccurately conclude that an intervention is ineffective, when its true effectiveness could be substantial if implemented as intended. Conversely, comprehensive FOI data could reveal that a reduced level of intervention is sufficient. Moreover, detailed FOI data provides valuable insights into the specific outcome measures influenced by the intervention, illuminating discrepancies between perceived and actual intervention effects.

Keywords:

Fidelity, intervention, children, parent-implemented, stuttering, fluency

Acknowledgements

This journey has been an adventure, to say the least. It has taken ten years, filled with challenges, growth, and countless memorable moments. Throughout this time, I have had the pleasure of working with some fantastic people and would not have finished without their help and support.

First and foremost, I want to express my deepest gratitude to my supervisor, Jóhanna, whose guidance, expertise, and unwavering support have been instrumental in this journey. Her patience and encouragement have been invaluable, and I am profoundly grateful for her mentorship.

I also want to thank Roger Ingham for welcoming my family and me with open arms during our time in the USA. His supervision, coupled with the warm introduction to Santa Barbara, created special memories that our family will always cherish.

A heartfelt thanks to Kate Crowe, whose tireless efforts and unwavering support have been the true driving force behind my progress. Kate has been the cornerstone of my success, tirelessly working behind the scenes to ensure I stayed on track. Her relentless dedication, practical assistance, and emotional support have been invaluable. Her ability to manage everything with such grace and competence has truly been remarkable, and I am eternally grateful for her consistent support.

I am deeply grateful to Þorlákur, whose expertise in data and kind, calm demeanor made every complex task manageable. His contributions have been crucial to the completion of this dissertation.

I must also extend my deepest thanks to Þórarinn, whose constant mantra of "just finish" kept me focused and motivated. His encouragement and aim to see me graduate pushed me to persevere through the toughest times.

Lastly, I want to thank my family for their endless support and love. This journey would not have been possible without them by my side, offering constant encouragement and understanding.

To all of you, this accomplishment is as much yours as it is mine. Thank you for being part of this remarkable journey.

I am incredibly happy to have reached this milestone, although I must admit, it is not an experience I would likely choose to undergo again.

Contents

List of Abbreviations

CWS	Child/Children who Stutters
DCM	Demands and Capacities Model
FOI	Fidelity of Implementation
OASES-S	Overall Assessment of the Speaker’s Experience of Stuttering – School Age
SLPs	Speech-Language Pathologists
SPM	Syllables Per Minute
STOC	Stuttering Treatment for Older Children
%SS	Percentage of Syllables Stuttered

List of Figures

List of Tables

List of Original Papers

This thesis is based on the following original publications, which are referred to in the text by their Roman numerals (I - IV):

Paper I

Bergþórsdóttir, I. Ö., & Ingham, R. J. (2016). Putting the cart before the horse: A cost effectiveness analysis of treatments for stuttering in young children requires evidence that the treatments analysed were effective. *Journal of Communication Disorders, 65*, 65-67. https://doi.org/10.1016/j.jcomdis.2016.04.006

Paper II

Bergþórsdóttir, Í. Ö., Crowe, K., & Einarsdóttir, J. T. (2022). Implementation fidelity in parent-implemented interventions for stuttering. *Clinical Linguistics & Phonetics, 36*(10), 904–927. https://doi.org/10.1080/02699206.2021.1965659

Paper III

Einarsdóttir, J. T., Karlsson, T., Bergþórsdóttir, Í. Ö., Halldórsdóttir, K.L., Crowe, K. (2025). A pilot study of stuttering treatment for older children (STOC). *Journal of Fluency Disorders, 83*, 106102. https://doi.org/10.1016/j.jfludis.2025.106102

Paper IV

Bergþórsdóttir, I. Ö., Crowe, K., Einarsdóttir, J. T., & Karlsson, Þ. (2024). The effect of fidelity in a pilot study of stuttering treatment for older children. *Manuscript submitted for publication.*

Original publications authored during the candidate's doctoral studies that further demonstrate the candidates' work on fidelity and intervention. These cannot form part of the thesis as they are published in Icelandic:

Appendix A

Sigfúsdóttir, S. A., Einarsdóttir, J. T., Karlsson, Þ., & Bergþórsdóttir, I. Ö. (2020). Orðaforðakennsla með sögulestri fyrir börn með málþroskaröskun [Vocabulary intervention for children with Developmental language Disorder]. Netla – *Veftímarit um uppeldi og menntun*. Menntavísindasvið Háskóla Íslands. https://doi.org/10.24270/netla.2020.4

Appendix B

Einarsdóttir, J. T., Agnarsdóttir, E., & Bergþórsdóttir, I. Ö. (2022). Tvítyngd börn, orðaforði og sögulestur [Bilingual children, vocabulary and storybook reading]. Einarsdóttir, J (Eds). *Leikandinn: Greinar um menntun ungra barna* (pp. 137-163). Háskólaútgáfan.

All papers are reprinted by kind permission of the publishers.

Declaration of Contribution

Paper I. Putting the cart before the horse: A cost effectiveness analysis of treatments for stuttering in young children requires evidence that the treatments analysed were effective (published in 2016, Journal of Communication Disorders)

This paper examines the validity of the findings by de Sonneville-Koedoot, Bouwmans, Franken and Stolk (2015), which focused on the cost-effectiveness of two treatment programs for young children who stutter. My contribution to this paper involved conducting the necessary calculations and thoroughly reviewing the studies referenced in the findings as well as writing the draft of the paper. This paper serves as a foundational piece for future research, as it underscores the critical importance of understanding what is being implemented during interventions.

Paper II. Implementation fidelity in parent-implemented intervention for stuttering (published in 2022, Clinical Linguistics and Phonetics)

My contribution to this paper encompassed several key areas. I was responsible for planning the overall research, which included developing the framework and approach for the study. I planned and trialled the systematic literature search and then led the data collection process. I reviewed each paper that emerged from the search process for whether it met the inclusion criteria for the review, extracted relevant data and coded this for papers that met the inclusion criteria and coordinated others to calculate inter-rater reliability. I was responsible for all calculations and analyses of the data, which involved processing and interpreting data related to FOI across the reviewed studies, ensuring that the results were accurately represented and aligned with the research objectives. These calculations were crucial in identifying trends, differences, and key findings, which ultimately informed the discussion and conclusions presented in the paper. As the lead author, I was responsible for planning and writing the first draft of the manuscript and integrating feedback from co-authors into subsequent drafts. After submission of the article, I was responsible for addressing reviewers' comments and writing the *response to reviewer's* document. I reviewed, edited, and approved pre-publication page proofs of the article.

Paper III. A pilot study of stuttering treatment for older children
(published in 2025, Journal of Fluency Disorders)

My contribution to this paper involved several key tasks. I planned the fidelity analysis, collected the necessary data, and made the calculations related to fidelity that are included in this paper. My analytical work included interpreting the data, identifying trends, and drawing meaningful conclusions, all of which were vital in framing the overall findings of the study. I played a central role in writing the paper, particularly sections related to the FOI data in the introduction, methods, results, and discussion. I worked as part of the authorship team to address reviewers' comments, revise the manuscript, and prepare the *response to reviewer's* document. In this, I took the lead on revisions related to FOI.

Paper IV. The effect of fidelity in a pilot study of stuttering treatment for older children (in submission, Journal of Evidence-Based Communication Assessment and Intervention)

In this paper, FOI data from the study described in Paper III were analysed in a more detailed way to examine the relationship between FOI and the intervention outcomes. By examining the FOI at a deeper level, I was able to explore patterns and correlations between the intervention implementation and the participant outcomes, providing valuable insights into the study's impact. I was the leader of conducting this study and writing the article. I planned the methodology and data analysis for this study and conducted the analyses (with the support of committee member Þorlákur Karlsson for the Pearson Product-Moment correlations analysis). I interpreted the results of analyses to answer the research questions posed in this study and created all the figures and tables in the article. As the lead writer, I planned and wrote the first draft of the manuscript. I integrated feedback from co-authors into subsequent drafts.

All of the above was done with the sound guidance of my supervisor and doctoral committee and in close collaboration with the co-authors of each paper.

1 Introduction

In recent years, evidence-based research has increasingly focused not only on determining the real-world effectiveness of interventions but also on understanding the underlying reasons for their efficacy in controlled settings. As a result, there has been progress in the development of guidelines governing the evaluation and execution of intervention fidelity in research of behavioural interventions. Various frameworks have emerged to address fidelity considerations, such as those proposed by An et al. (2020), Bellg et al. (2004), Century et al. (2010), and Lieberman-Betz (2015). Fidelity of implementation (FOI), a key construct in this context, refers to the assessment of whether interventions were carried out as intended, as defined by Perepletchikova and Kazdin (2005). Measuring FOI involves quantifying the disparity between the intended and actual implementation of interventions.

FOI denotes the extent to which an intervention adheres to its intended or planned execution (An et al., 2020; Dane & Schneider, 1998; Dusenbury et al., 2003; Perepletchikova & Kazdin, 2005; Prowse et al., 2015). When an intervention deviates from its intended implementation, establishing a causal link between the independent variable (the intervention) and changes in the dependent variable (stuttering behaviour) becomes challenging (Gresham et al., 1993). Inadequate fidelity in intervention research jeopardizes both internal and external validity, rendering causal inferences unreliable, as observed by Gresham et al. (1993) and Moncher and Prinz (1991). Neglecting FOI could lead to erroneous conclusions regarding the efficacy of individual interventions and hinder comparisons between different intervention approaches in clinical settings (Bergþórsdóttir & Ingham, 2016). Hence, FOI has emerged as a pivotal consideration in intervention study guidelines, such as the Single-Case Reporting Guideline in Behavioural Intervention, also known as SCRIBE (Tate et al., 2016), and in research quality assessments, exemplified by the Council for Exceptional Children standards for evidence-based practices in special education (CEC, 2014). In essence, failure to demonstrate adequate fidelity in clinical intervention research represents a significant problem that impedes the ability of researchers and practitioners to gauge intervention effectiveness and ascertain whether a particular study and intervention merits inclusion in the evidence base for practice (Hofslundsengen et al., 2022, Kaderavek & Justice, 2010; McCormack et al., 2017).

1.1 The construct of fidelity of implementation

Research on complex or multifaceted interventions tends to have poorer fidelity compared to studies focusing on simpler interventions, as complex intervention often involves various components and procedures that make it more challenging to have faithful implementation across different settings and practitioners (Hogue et al., 2006). Behavioural interventions frequently face difficulties in maintaining fidelity records, particularly when interventions are implemented outside clinical settings. This challenge arises from the intricate nature of these interventions, which often encompass multifaceted strategies and techniques that can pose challenges to consistent implementation across diverse settings and by different practitioners. Variations in practitioners' experience, training, and proficiency levels may result in inconsistencies in fidelity. Implementations of mental health and behavioural interventions often occur in non-clinical settings, where the unique needs and traits of clients being served can influence delivery methods, further exacerbating fidelity challenges (An et al., 2020; Schoenwald et al., 2011). Perepletchikova and Kazdin (2005) described the challenge of devising universally applicable FOI measures, stating that since interventions can differ in their components and the necessary qualifications for effective implementation, measures of fidelity might need to be customized for individual interventions.

Fidelity refers to the degree to which an intervention is delivered according to the prescribed protocol or guidelines. Establishing whether the intervention was faithfully executed allows researchers to determine whether observed outcomes can be attributed to the intervention itself (Ginsburg et al., 2021; Walton et al., 2020). Treatment fidelity involves continuously assessing, monitoring, and improving the reliability and internal validity of a study, including identifying errors of commission (adding unnecessary elements) and errors of omission (missing essential components) (Borrelli et al., 2005). While numerous terms and dimensions have been proposed for assessing FOI in intervention studies, the literature has coalesced around a cohesive understanding of its fundamental components, which encompass dosage, adherence, quality, and responsiveness (An et al., 2020). Building on the groundwork of the National Institutes for Health Behavior Change Consortium (NIH BCC) treatment fidelity framework (Bellg et al., 2004; Borrelli et al., 2005), a comprehensive five-domain model for evaluating treatment fidelity, this understanding meticulously considers aspects such as study design, training, delivery, receipt, and enactment. Despite the absence of commonly used guidelines or best practices for evaluating FOI, researchers acknowledge its pivotal role in shaping intervention outcomes (Hofslundsengen et al., 2022). Consequently, researchers must meticulously design intervention protocols that incorporate rigorous FOI monitoring (An et al., 2020; Carroll et al., 2007; Dusenbury et al., 2003; Hagermoser Sanetti, & Kratochwill, 2014; Ingersoll & Dvortcsak, 2006; Lieberman-Betz, 2015; Roberts & Kaiser, 2011; Walton et al. 2020). The NIH BCC treatment fidelity framework (Bellg et al., 2004; Borrelli et al., 2005) have been validated and updated (Borrelli et al., 2011) to provide a comprehensive framework for evaluating treatment fidelity (Johnson-Kozlow et al., 2008; Spillane et al., 2007; Wyatt et al., 2010). The five-domains in their model are design, training, delivery, receipt, and enactment. *Design* involves considering and reporting factors necessary for trial evaluation and replication, such as treatment content, dose, provider credentials, and theoretical frameworks. *Training* focuses on developing and standardizing training for providers, measuring skill acquisition, and maintaining skills over time. *Delivery* supports the intervention being delivered as intended, with mechanisms to assess provider adherence and measure nonspecific effects. *Receipt* confirms participants understand the intervention information. Finally, *enactment* monitors and improves patients' ability to apply cognitive strategies and behavioral skills in daily life.

The selected references represent a comprehensive body of literature that collectively underscores the importance of FOI in intervention research. By citing these authors, I aim to demonstrate the consensus among researchers regarding the critical role of FOI in shaping intervention outcomes. Each of these authors has contributed valuable insights into various aspects of FOI assessment, emphasizing its significance in ensuring the integrity and effectiveness of interventions. Therefore, by referring to these authors I seek to draw attention to the robust evidence base supporting the meticulous monitoring and evaluation of FOI in intervention studies. In the next section, the four elements of FOI will be defined, drawing on literature that elucidates the critical components and measurement techniques associated with FOI.

1.1.1 Dosage

In behavioural interventions, dosage plays a critical role, encompassing the frequency and duration of intervention sessions or activities (An et al., 2020; Hofslundsengen et al., 2022). Dosage involves tracking how often and for how long participants engage in the intervention, ensuring they receive the prescribed amount as outlined in the study protocol (An et al., 2020; Bellg et al., 2004). This concept is essential for maintaining uniformity across different conditions, especially when multiple behavioural targets are involved. Consistency in dosage maintains the integrity and comparability of study outcomes. However, dosage in behavioural contexts extends beyond frequency and duration of intervention. In this context it can also include the amount of program content that the participants receive (Slaughter et al., 2015). This broader view of dosage encompasses multiple dimensions, including self-reports by providers for all lessons, objective assessments based on observations, and attendance records for each participant. This multifaceted approach provides a comprehensive understanding of the intervention's reach and impact (Dusenbury et al., 2003; Walton et al., 2020).

Detailed assessment of dosage captures various aspects such as how often a target is addressed, the duration of each session, the frequency of intervention delivery per week or month, the time span over which the intervention occurred, and the total time spent in intervention over the entire intervention period (Kaderavek & Justice, 2010). This comprehensive approach fosters transparency and rigor in evaluating intervention effectiveness (Perepletchikova et al., 2007; Walton et al., 2020). Overall, by addressing these multiple facets of dosage, researchers can promote the effective implementation and evaluation of behavioural interventions.

1.1.2 Adherence

Adherence in intervention delivery is a fundamental aspect, evaluating the extent to which intervention providers, such as clinicians, educators, or parents, adhere to prescribed procedures or techniques during intervention implementation (An et al., 2020). It entails closely monitoring whether interventionists follow the intervention protocol, including specific instructions, techniques, or strategies (Bellg et al., 2004). Ensuring adherence is crucial for maintaining consistency in intervention delivery across different providers and settings. It is essential to prevent contamination across interventions and control conditions, particularly when administered by the same provider. Efforts must be made to maintain fidelity to the intended content of intervention, thereby enhancing the reliability and validity of study outcomes (Bellg et al., 2004).

Adherence extends beyond the consistency of intervention delivery to include the degree of implementation of specified program components as outlined in program manuals. This involves ensuring that intervention activities and methods are consistently implemented according to the program's written guidelines (Dane & Schneider, 1998; Ginsburg et al., 2021). Treatment fidelity encompasses not only the correct implementation of an intervention but also the identification of errors of commission and omission, which can significantly impact dissemination of the intervention (Borrelli et al., 2005). Assessing adherence can employ various approaches, such as self- or other-completed checklists to track implementation accuracy, direct observation of intervention sessions, and measuring implementation as a dependent variable. These methods offer insights into how well the implementer follow the prescribed procedures and support the ongoing monitoring of intervention fidelity (Lieberman-Betz, 2015). Borrelli (2011) highlighted that digital recorders and other electronic monitoring solutions can be used to give specific feedback to providers, promote standardization, and store data efficiently. Furthermore, adherence involves aligning the implementation of specific activities and methods with the program's written instructions. It specifies the implementation strategies and evaluates the extent to which these strategies are executed as intended (Slaughter et al., 2015). Practically, assessing adherence focuses on whether the implementer adheres to the correct steps during implementation, ensuring that interventions are delivered consistently and according to protocol (Kaderavek & Justice, 2010). Both the manner in which adherence is monitored (e.g., direct observation vs. self-assessment) and the rigour with which this is implemented impact the accuracy of adherence measurement.

Overall, assessing adherence can involve using one or more strategies, including direct observation, self-report measures, and checklists. Additional strategies include using digital tools for real-time monitoring and feedback (Bellg et al., 2004; Borrelli et al., 2005; Borrelli, 2011). These tools can track adherence to protocols, monitor the delivery of interventions, and provide insights into the effectiveness of different strategies. It is essential to assess adherence to support the effective delivery of interventions and the validity of study findings

(Perepletchikova et al., 2007). This comprehensive approach to adherence helps maintain the integrity and efficacy of behavioural interventions.

1.1.3 Quality

Quality in intervention delivery encompasses the proficiency with which intervention components are implemented, emphasizing the competence and skill of the implementer (An et al., 2020). In previous research quality has been described in terms of how well the implementer delivers the intervention, encompassing their ability to communicate effectively, establish rapport with participants, and deliver intervention content with clarity and accuracy (Bellg et al., 2004; Ginsburg et al., 2021). As stated by Bellg et al. (2004), it is important that training is conducted uniformly across different providers, that providers are trained to well-defined performance criteria, that their skills do not diminish over time, and that the delivery of treatment is consistently monitored. This comprehensive evaluation verifies that the implementer is proficient in delivering the intervention and that participants receive high-quality services (Lieberman-Betz, 2015).

Quality of delivery encompasses qualitative aspects of program delivery beyond the mere implementation of prescribed content. This includes factors such as the enthusiasm of the implementor, the overall effectiveness of the session, and the implementor’s attitudes towards the program. These elements significantly influence the overall effectiveness of the intervention and the engagement of participants (Dane & Schneider, 1998). Additionally, assessing the quality of delivery involves evaluating how closely a provider aligns with the theoretical ideal in delivering program content (Bellg et al., 2004; Borrelli et al., 2005; Borrelli, 2011). This validates that the implementor meets established standards for intervention delivery, facilitating ongoing monitoring of intervention fidelity (Dusenbury et al., 2003). This involves procedures such as rater training, assessing inter-rater reliability, and controlling for measure reactivity, overall assessing the effectiveness of the training provided to the individual delivering the intervention (Bellg et al., 2004). These measures help maintain the integrity and validity of intervention fidelity assessments, ensuring interventions are implemented as intended (Perepletchikova et al., 2007; Walton 2020). Overall, evaluating the quality of intervention delivery is crucial for ensuring the effectiveness and fidelity of interventions. By assessing implementors’ proficiency to program guidelines, researchers can enhance the validity and reliability of study findings and improve participant outcomes.

1.1.4 Responsiveness

Responsiveness in intervention evaluation focuses on participants' engagement, receptivity, and responsiveness to the intervention (An et al., 2020). It entails assessing how participants interact with and actively participate in intervention activities, including their level of comprehension, engagement, cooperation, and motivation (Bellg et al., 2004). One aspect of responsiveness is ensuring that participants comprehend intervention information, particularly in cases where cognitive abilities are compromised, literacy/education levels are low, or language proficiency is limited. It involves confirming participants' ability to apply cognitive skills taught in the intervention. Additionally, it involves confirming participants' ability to utilize the behavioural skills taught in the intervention (Bellg et al., 2004). Assessing participant responsiveness also entails evaluating participant reactions to program sessions, which may encompass indicators such as levels of involvement and enthusiasm. This assessment helps gauge the degree of participant engagement and involvement in program activities and content (Dane & Schneider, 1998; Dusenbury et al., 2003).

Participant responsiveness is measured based on reports or observations of participant engagement during training or intervention sessions. This may include implementors' satisfaction reports obtained through questionnaires or interviews, and observations or reports of the participants engagement with the intervention (Lieberman-Betz, 2015). The extent to which participants engage with and are involved in program activities and content is essential for the success of the intervention. It also considers the involvement of intervention providers in developing, evaluating, or being receptive to implementation strategies (An et al., 2020; Slaughter et al., 2015). By evaluating participant responsiveness, researchers can assess the degree to which participants are actively engaged in the intervention, identify areas for improvement, and tailor intervention strategies to better meet participants' needs. This comprehensive evaluation enhances the effectiveness and relevance of interventions, ultimately improving outcomes for participants.

1.1.5 Measurements

Various fidelity evaluation methods are used in research studies to confirm that interventions are implemented as intended and to assess the integrity of the intervention protocol. These methods are crucial for verifying that the intervention is delivered consistently and accurately, thereby enhancing the reliability of research findings. Depending on context, the reporting can vary depending on who is evaluating the intervention. This variability allows the assessment to be tailored to specific needs and perspectives, enhancing the overall understanding of intervention fidelity. Each method offers unique advantages and contributes to a comprehensive assessment of fidelity in intervention research.

Researchers and/or trained observers: Direct observation serves as a cornerstone method in fidelity evaluation, where researchers or trained observers actively supervise intervention sessions in real-time. Additionally, having a second observer present during intervention sessions enables real-time assessment of fidelity. The second observer can independently assess the delivery of the intervention and compare their observations with those of the primary observer. This method helps validate fidelity assessments and reduces the potential for bias in the evaluation process (Tong et al., 2019). Using standardized checklists or rating scales, they document the presence or absence of key fidelity components. This meticulous approach provides valuable insights into the actual delivery of the intervention, enabling immediate feedback to the intervention provider and fostering continuous improvement (An et al., 2020). This structured approach supports all critical aspects of the intervention being monitored and recorded systematically, as well acting to identify errors of commission and omission (Bellg et al., 2004; Borrelli et al., 2005; Borrelli, 2011). Similarly, audio or video recordings of intervention sessions are frequently utilized for later review and fidelity assessment. Researchers analyse these recordings to evaluate adherence to the intervention protocol, identify areas for improvement, and provide feedback to the intervention provider. The assessment and monitoring of fidelity is important for accurately evaluating whether or not the intervention is being adequately delivered. This method allows for a detailed examination of intervention delivery and facilitates the identification of specific areas for training and/or support (An et al., 2020: Boyle, 2015; Ginsburg et al., 2021). Documentation review involves examining session notes, intervention logs, or intervention manuals to assess fidelity to the intervention protocol. This method provides valuable insights into the delivery of intervention components and any modifications made during implementation (Ginsburg et al., 2021). It allows researchers to track changes in the intervention over time and maintains consistency in implementation across different settings or different providers (Craig et al., 2009; Walton et al., 2020). Additionally, as an extension of fidelity investigations, researchers may also consider approaches used in knowledge translation, implementation science, and/or social validity investigations to better understand the challenges faced in achieving fidelity from the user perspective. For example, questionnaires could be administered to intervention providers, participants, or both to gather feedback on their experiences with the intervention. By collecting self-reported data, researchers can gain insights into the factors that may impact fidelity and identify areas for improvement (Guitar, 2014). Structured interviews with intervention providers or participants offer another avenue for gathering feedback on intervention implementation. Researchers conduct these interviews to explore perceived adherence to the protocol, barriers to implementation, and suggestions for improvement. This qualitative approach adds depth to the fidelity assessment process by capturing the perspectives and experiences of those directly involved in the intervention (Walton et al., 2020).

Intervention provider: Self-report measures are another common approach, wherein intervention providers report on their adherence to the intervention protocol (An et al., 2020). They document their implementation practices, challenges encountered, and deviations from the protocol. While self-report measures offer valuable information, they may be subject to bias and may not always accurately reflect the actual implementation of the intervention (Yaruss & Quesal, 2016). For example, in school-based prevention programs, a teacher might report higher adherence to a program's protocol to avoid appearing non-compliant or understate deviations to seem more skilled (Dane & Schneider, 1998). This bias can lead to an overestimated sense of program fidelity, compromising the validity of research findings. Dane and Schneider (1998) highlight that fidelity ratings provided by trained observers show stronger effects than those from implementer self-reports, suggesting that self-reports are commonly affected by social desirability and are less reliable than objective measures like direct observation. Overcoming these challenges requires training, clear guidelines, support, and thorough documentation.

In summary, fidelity evaluation methods are essential for accurate and consistent implementation of interventions in research. By employing a combination of direct observation, self-report measures, audio or video recordings, structured interviews, and documentation review, researchers can comprehensively assess fidelity and make informed decisions to optimize intervention delivery (Yaruss & Quesal, 2004). Each method offers unique advantages and contributes to a comprehensive assessment of fidelity in intervention research.

1.1.6 FOI definitions used in this thesis

In this thesis four FOI components were considered: dosage, adherence, quality, and responsiveness. Use of these four components provides a straightforward and focused framework by concentrating on four essential components. The following definitions of the FOI components will be used. *Dosage* refers to the frequency of session delivery, indicated by the number of sessions attended, session duration, and duration of the entire intervention (An et al., 2020; Bellg et al., 2004; Dane & Schneider, 1998; Dusenbury et al., 2003; Kaderavek & Justice, 2010; Lieberman-Betz, 2015; Slaughter et al., 2015). Typically, dosage is measured through session attendance logs and/or session recordings. *Adherence* pertains to the accuracy of intervention delivery and is often assessed using checklists, automatic recording devices, and/or independent observer judgments (An et al., 2020; Bellg et al., 2004; Dane & Schneider, 1998; Dusenbury et al., 2003; Kaderavek & Justice, 2010; Lieberman-Betz, 2015; Perepletchikova et al., 2007; Walton et al., 2020). *Quality* denotes the effectiveness of intervention delivery and describes the skilfulness with which the intervention was delivered and the ability of the interventionist to engage the participant (An et al., 2020; Bellg et al., 2004; Dane & Schneider, 1998; Dusenbury et al., 2003; Gearing et al., 2011; Kaderavek & Justice, 2010; Lieberman-Betz, 2015; Perepletchikova et al., 2007). Quality is often assessed using similar methods as adherence, such as checklists, automatic recording devices, and/or independent observer judgments. To illustrate the distinction between adherence and quality, consider an individual who consistently performs a prescribed exercise routine (adherence) but executes the exercises with poor technique and minimal effort (quality). While adherence verifies that the intervention is being followed as planned, quality reflects the effectiveness and skillfulness with which the intervention is delivered. High adherence with low quality may not yield the desired outcomes, underscoring the importance of both components in evaluating intervention success. While some have suggested that quality also includes the interventionist's adherence to the intervention protocol (e.g., Gearing et al., 2011), in this study the accuracy of adherence to protocol was not included as a component of quality, as this is considered as the FOI component of adherence. Lastly, *responsiveness* reflects participant involvement and engagement in the intervention, which includes gauging their comprehension and satisfaction with the intervention. By using satisfaction surveys and direct observation of participant behavior, it is possible to assess how well participants understand and engage with the intervention (An et al., 2020; Bellg et al., 2004; Century et al., 2010; Dane & Schneider, 1998; Dusenbury et al., 2003; Lieberman-Betz, 2015; Slaughter et al., 2015).

Quality and responsiveness are closely related components that interact and on occasions may overlap. In this thesis, quality predominantly focused on the behaviour of the interventionist (i.e., clinician or parent), while responsiveness focused predominantly on the behaviour of the participant (i.e., the child).

The rationale for choosing these components, above others, is that concentrating on dosage, adherence, quality, and responsiveness, allows for a streamlined evaluation process. This can make it easier to collect and analyze FOI data, leading to more efficient and actionable findings. Use of these FOI components is well supported by contemporary research and have been widely used in recent studies (An et al., 2020; Bellg et al., 2004; Century et al., 2010; Dane & Schneider, 1998; Dusenbury et al., 2003; Kaderavek & Justice, 2010; Lieberman-Betz, 2015; Perepletchikova et al., 2007; Slaughter et al., 2015; Walton et al., 2020). This alignment with current scientific practices lends credibility to the approach, grounding it in established methodologies. While the NIH BCC treatment fidelity framework (Bellg et al., 2004; Borrelli et al., 2005) offers a comprehensive evaluation by including design, training, receipt, and enactment of treatment skills, the chosen FOI components specifically target the core aspects of implementation rather than the intervention process more generally. This focused approach provides a thorough yet streamlined evaluation of intervention fidelity and effectiveness, complementing the broader scope of the NIH BCC treatment fidelity framework.

1.2 Stuttering

Developmental stuttering, also referred to as childhood-onset fluency disorder (DSM 5; American Psychiatric Association, 2013), is categorized as a neurodevelopmental disorder according to the International Statistical Classification of Diseases and Related Health Problems (11th ed.; World Health Organization, 2018). Stuttering is characterized by a persistent, frequent, or pervasive disruption of the rhythmic flow of speech that emerges during childhood development. In the context of stuttering, fluency refers to the smooth, uninterrupted flow of speech. This disruption is beyond the normal variation expected for a person's age and intellectual functioning, leading to reduced speech intelligibility and significantly impacting communication (World Health Organization, 2018).

These disruptions can range from repetitions of sounds, syllables, and/or words to prolongations of sounds and/or complete blocks where the speaker is unable to produce any sound. These disruptions can interfere with communication and may be accompanied by physical tension or struggle as the individual attempts to speak. The severity of stuttering can vary widely among individuals and can also fluctuate in different situations (Bloodstein et al., 2021; Constantino et al., 2016; Guitar 2014). While stuttering typically emerges during childhood, it can persist into adulthood and significantly impact various aspects of a person's life (Einarsdóttir et al., 2024; Yairi & Ambrose, 2013). The exact aetiology of stuttering remains elusive, but it is widely recognized as a multifactorial disorder influenced by genetics and/or neurological disorder (Ingham et al., 2018; Smith & Weber, 2016). Concordance rates among twins provide evidence supporting the genetic basis of stuttering (Rautakoski et al., 2012). Subsequent advanced aggregation studies and segregation analyses yielded data consistent with established genetic models. Recent linkage and association studies have started to identify specific genes that contribute to the disorder and recent studies have shown promising results and that genes with roles in structural organization and various neural processes might play a role in developmental stuttering risk (Polikowsky et al., 2022). Neuroimaging research has provided insights into the neurobiological basis of stuttering, revealing differences in brain structure and function, particularly in regions associated with speech production and motor control (Ingham et al., 2018; Neef et al., 2015). Understanding these classifications is essential for tailoring effective interventions for individuals affected by stuttering.

Developmental stuttering is the most common form of stuttering and typically emerges during early childhood, often between the ages of 2 and 5. This type of stuttering occurs during the developmental stages of speech and language acquisition when children are learning to formulate and produce sounds, words, and sentences (Hofslundsengen et al., 2022; Reilly et al., 2009; Yairi & Ambrose, 2013). Developmental stuttering may manifest as repetitions of sounds, syllables, or words, prolongations of sounds, or blocks where the flow of speech is interrupted. While the exact causes of developmental stuttering are not fully understood, it is believed to result from a combination of genetic, neurological, and environmental factors (Smith & Weber, 2016; Yairi & Ambrose, 2013). The onset of developmental stuttering is typically around the age of three years and is estimated to effect around 5%-8% of preschool children, with males being more likely to experience stuttering than females (Bloodstein et al., 2021; Yairi & Ambrose, 2013). Spontaneous recovery, recovery without therapy, occurs in around 60-80% of cases depending on the criteria used to define recovery (Einarsdóttir et al., 2020; Yairi & Ambrose, 2005). As children grow older, the likelihood of spontaneous recovery decreases, resulting in the need to start therapy early (Einarsdóttir et al., 2024; Yairi & Ambrose, 2005), with spontaneous recovery rarely seen by the time children are in their school years (Einarsdóttir et al., 2024). The incidence of stuttering is estimated to be around 4–5% in children, while the prevalence rate globally is approximately 1% (Bloodstein et al., 2021). In adulthood, the persistent developmental stuttering is observed more often in males than females, with a male-to-female ratio of 4:1 (Bloodstein et al., 2021).

Two types of stuttering typically have their onset during adulthood. Neurogenic stuttering typically develops later in life, often as a result of damage to areas of the brain responsible for speech and language processing (Neef at al., 2015), particularly impacting the brain's ability to control speech production (Junuzovic-Zunic et al., 2021). This type of stuttering can occur following a sudden injury to the brain, such as in stroke and traumatic brain injury, progressive neurological disorders, such as Parkinson's disease and multiple sclerosis, or the use of certain medications. Neurogenic stuttering is a multifaceted disorder, the underlying pathophysiological mechanisms of which are not fully understood. Functional stuttering differs from developmental or neurogenic stuttering through its adult onset, occurring without a prior history of speech difficulties. It is often characterized by extremes in variability or consistency in repetitions (Baker et al., 2021). Case studies indicate that functional dysfluency may emerge following stressful life events, particularly those involving conflict or challenges in expressing negative emotions with close individuals or authority figures (Baker et al., 2021). Differential diagnosis of neurogenic and functional stuttering is challenging for clinicians, especially as this determines the treatment options and prognosis for recovery. However, treatment typically involves collaboration between speech therapists and physicians, most commonly neurologists, and sometimes psychologists (Junuzovic-Zunic et al., 2021).

Stuttering can have significant, negative, effects on many different aspects of an individual's life. It can impair communication skills, leading to frustration and difficulties in expressing thoughts and ideas clearly (Iverach et al., 2009). Children, adolescents, and adults who stutter often encounter negative stereotypes and reactions from listeners (Snyder, 2001). These issues become more pronounced during adolescence, adversely affecting self-esteem, increasing anxiety levels, and harming social relationships and academic performance (Blood & Blood, 2004). The interruptions in speech flow caused by stuttering can impede effective communication, resulting in misunderstandings and hindering the conveyance of messages. Individuals who stutter often experience challenges in social interaction (Gunn et al., 2019). Many individuals form negative attitudes towards speaking and often face avoidance, struggle, or anxiety in speech situations (Iverach et al., 2009; Menzies et al., 2019). These challenges can lead to feelings of helplessness, shame, and embarrassment, and can generate expectations of social harm. Consequently, they may negatively impact one's occupational and educational achievements and overall quality of life (Yaruss, 2001). As a result, adults who stutter might have a higher risk of developing psychological, emotional, and behavioural difficulties (Craig, 2003).

CWS are often perceived as less popular and less likely to be seen as leaders compared to their peers who do not stutter (Davis et al., 2002). Childhood bullying has been linked to increased anxiety later in life (Gladstone et al., 2006; Hawker & Boulton, 2000; McCabe et al., 2010), and studies indicate that CWS report experiencing significantly higher rates of bullying than those who do not. Over time, the cumulative effect of these experiences may contribute to the development of depression or other mental health concerns (Bloodstein et al., 2021; Sizer & Sizer, 2023). The fear of stuttering in public settings can lead to social anxiety and avoidance of speaking situations (Craig et al., 2002; Iverach & Rapee, 2014). People who stutter may avoid certain social gatherings, public speaking engagements, or situations where they anticipate judgment or negative reactions regarding their speech (Helgadóttir et al., 2014). The emotional toll of stuttering can be substantial. Negative reactions from others, such as ridicule or impatience, can exacerbate these emotions, leading to increased social isolation and diminished self-confidence (Gunn et al., 2019). Studies have even associated stuttering with reduced earnings (Gerlach et al., 2018).

1.3 Stuttering intervention for different age groups

Interventions tailored to different age groups aim to address the unique needs and challenges associated with developmental stuttering at various stages and ages. In this context, interventions for preschool-aged children, school-aged children, and adults who stutter adopt distinct approaches to promote fluency, communication confidence, and effective coping strategies (Baxter et al., 2016; Guitar, 2014).

There is evidence supporting the benefits of early intervention, in addressing stuttering, particularly in the preschool years. Early identification and intervention have been associated with improved long-term outcomes, including enhanced speech fluency and reduced psychosocial impact (Baxter et al., 2016; Bothe et al., 2006; Brignell et al., 2021; Hofslundsengen et al., 2022). These interventions typically focus on early identification of stuttering behaviours and implementation of strategies to promote fluent speech in naturalistic settings (Hofslundsengen et al., 2022; Reilly et al., 2009). By intervening during this critical period of language development, clinicians and parents can mitigate the long-term impact of stuttering on a child's communication skills and overall well-being (Onslow & Packman, 1999; Roberts & Kaiser, 2011). However, several challenges and limitations accompany early intervention efforts, such as the potential difficulty in accurately distinguishing between typical disfluencies and early signs of persistent stuttering, especially in young children who are still developing their speech and language skills (Yairi & Ambrose, 2013). Many children who exhibit early signs of stuttering naturally outgrow their disfluencies without formal intervention (Einarsdóttir et al., 2020). The recovery rate for stuttering is high in the preschool years, but the chances of recovery diminish as children get older. Since not all children outgrow stuttering in their early years, it is essential to develop effective strategies to improve fluency in school-aged CWS (Bothe et al., 2006; Einarsdóttir, et al., 2020). This natural recovery complicates the determination of which children require immediate intervention, leading to the risk of overtreating children who may not develop persistent stuttering (Bothe et al., 2006).

With greater cognitive abilities and comprehension, school-aged children can actively engage in therapy sessions, set goals, and participate in self-monitoring of their speech (Bothe et al., 2006). School-aged children have typically developed more sophisticated language and communication abilities compared to preschoolers (Bloodstein et al., 2021). This developmental milestone allows speech-language pathologists (SLPs) to employ a greater array of intervention techniques, including cognitive-behavioural strategies and speech modification approaches, which are instrumental in enhancing fluency and communication effectiveness. School-aged children also benefit from increased opportunities for social interaction with peers, both within and outside of therapy settings (Baxter et al., 2016; Bothe et al., 2006; Einarsdóttir, et al., 2020). Stuttering interventions for school-aged children (6 to 12 years of age) are uniquely positioned to integrate academic content and classroom activities, promoting the generalization of fluency techniques into real-life situations (Guitar, 2014). The school system for compulsory schooling in Iceland is structured into three levels: the first level for ages 5-9, the middle level for ages 9-12, and the upper level for ages 12-15. Collaboration between SLPs and educators enables the implementation of strategies that support fluency in academic tasks and social interactions within the school environment. By comprehensively understanding the child's speech patterns, environmental influences, and social dynamics, therapists can tailor intervention plans to address specific needs and challenges (Baxter et al., 2016; Blood & Blood, 2004; Bothe et al., 2006; Guitar, 2014). These personalized plans often encompass a combination of direct speech therapy, classroom accommodations, and collaborative efforts with parents and teachers to foster fluency and confidence in communication (Baxter et al., 2016; Tubele et al., 2022). The primary objective of stuttering therapy is to reduce or eliminate stuttering. If complete elimination is not achievable, the focus shifts to teaching individuals effective strategies to manage their stuttering (Blomgren, 2013; Craig et al., 2009).

Finally, interventions for adults with developmental stuttering, prioritize enhancing communication confidence and fostering effective coping strategies. These interventions acknowledge the psychosocial impact of stuttering on adults by aiming to improve their quality of life through tailored therapeutic approaches (Laiho et al., 2022). By addressing both the physical aspects of stuttering and the emotional challenges associated, interventions for adults facilitate meaningful improvements in communication and overall well-being (Ingham et al., 2012; Yaruss & Quesal, 2016). By adopting age-appropriate strategies and addressing the multifaceted nature of stuttering, these interventions strive to promote fluent speech, enhance communication skills, and empower individuals to effectively manage their stuttering.

1.4 Stuttering intervention approaches

Two primary intervention approaches have been used for addressing stuttering, especially in preschool-aged children: direct and indirect intervention approaches (for overview see Sidavi and Fabus, 2010). Direct approaches involve providing the child with explicit instructions and feedback to improve fluency. Indirect approaches focus on modifying the child’s environment to create a fluency-enhancing atmosphere. In direct approaches parents and clinicians give immediate feedback during conversations, offering praise for fluent speech and correcting stuttered speech in a supportive manner. The goal of this approach is for the child to monitor and improve their speech fluency through clear instructions and feedback. One prominent example of a direct intervention approach is the Lidcombe Program (Johnson et al., 2024; Onslow et al., 2003). In the Lidcombe Program, verbal contingencies are delivered in a positive manner to support fluent speech being a positive experience for the child, for example, saying, “That was really smooth talking”. Verbal contingencies for unambiguous stuttering are also used in the Lidcombe program are present, but used sparingly and carefully (Onslow et al., 2020). In contrast, in indirect approaches rather than giving direct feedback on stuttered speech, parents adjust the way they communicate to reduce pressure and demands on the child (Franken et al., 2005). These adjustments may include slowing down their own speech, asking simpler questions, and reducing the pressure for immediate answers. The goal is to help the child feel more at ease and reduce the likelihood of stuttering episodes. The Demands and Capacities Model (DCM) is an example of an indirect intervention (de Sonneville-Koedoot et al., 2015; Franken et al., 2005).

According to a meta-analysis by Baxter et al. (2016), both direct and indirect interventions can be effective in managing stuttering in preschool-aged children. The study found that the effectiveness of these approaches can vary based on individual needs and circumstances, highlighting the importance of tailored therapeutic approaches. Additionally, the meta-analysis by Mallick et al. (2021) explored school-aged stuttering interventions and emphasized the need for evidence-based practices that consider the diverse contexts and quality of studies. This review further supports the importance of personalized intervention plans to address the specific needs of each child, ultimately aiming to improve their fluency and confidence in communication. Direct interventions, particularly interventions that involve operant conditioning are the focus of this thesis due to their robust empirical support, structured approach, adaptability to individual needs, and their holistic impact on both speech fluency and the psychosocial aspects of stuttering.

1.4.1 Operant conditioning in stuttering intervention

Operant conditioning, as defined by Skinner (1965), was originally a behavioural theory that focuses on modifying behaviour through rewards and punishments, described as a learning process in which behaviours are influenced by the consequences that follow them. The theory is based on the idea that behaviours followed by positive outcomes are likely to be repeated, while behaviours followed by negative outcomes are less likely to occur again (Skinner, 1965). Operant conditioning also describes the relationship between behaviour and environment (Poling et al., 2002). Events or conditions in the environment that can influence the expression of a behaviour. These events and conditions are called antecedents. The Lidcombe Program and the Response Contingent Time-Out are both examples of direct approaches using operant conditioning, where speech behaviours are either shaped through reinforcement (praise for fluent speech) or correction, mild punishment (brief time-outs for stuttered speech) with the goal to increase fluent speech by rewarding desired behaviours and discouraging stuttering through consequences (Hewat et al., 2006; Johnson et al., 2024). While in general, positive approaches are recommended as first options in behavioural interventions, time-out interventions have consistently been found to be effective in the treatment of stuttering in childhood (for review see Brignell et al., 2021).

Response Contingent Time-Out is a behavioural intervention technique commonly employed in stuttering therapy (Bloodstein et al., 2021). *Time-out*, also called *time-out from reinforcement*, entails an immediate interruption or time-out from speaking following the occurrence of a stuttered speech event. The underlying objective of Response Contingent Time-Out is to diminish the frequency and severity of stuttering behaviours by administering a consequence for stuttering occurrences (Hewat et al., 2006; James et al., 1989). During Response Contingent Time-Out sessions, when a child who stutters experiences a stuttered speech event, the therapist or parent implements a brief pause or interruption in the conversation. This interruption functions as a time-out period immediately following the occurrence of stuttering. During the time-out there are usually no specifications on behaviour other than that there is a pause from both participants speaking in the conversation. Typically lasting a few seconds, the time-out is consistently applied following each instance of stuttering to effectively associate the consequence with the stuttering behaviour (Hewat et al., 2006; James et al., 1989). Response Contingent Time-Out primarily aims to provide a prompt and consistent consequence for stuttering. By interrupting speech flow post-stuttering, Response Contingent Time-Out seeks to deter the recurrence of stuttering behaviours over time (Bloodstein et al., 2021). The technique is grounded in operant conditioning principles, specifically the concept of punishment, where stuttering is viewed as an undesirable behaviour (Hewat et al., 2006; James et al., 1989). Through repeated application, the child is anticipated to decrease stuttering to avoid the time-out period. Response Contingent Time-Out may be incorporated into a comprehensive stuttering therapy regimen alongside other techniques such as speech restructuring (modifying speech patterns to improve fluency), fluency shaping (teaching smooth, fluent speech through controlled breathing and gentle onset of speech), and cognitive-behavioural (addressing negative thoughts and behaviors associated with stuttering) strategies (Bloodstein et al., 2021). It is often customized to suit the individual needs of the child and may be complemented with positive reinforcement for fluent speech. Hewat et al. (2006) investigated Response Contingent Time-Out in a study with 22 adults and adolescents who stuttered, with over half of the participants reducing the percentage of stuttered syllables (%SS) by more than 50%. The majority of participants expressed satisfaction with the intervention, noting its ease of use and effectiveness compared to interventions centred on speech restructuring, such as prolonging speech. Despite this, the study had several limitations, including the need to identify which clients benefit most from the program and the extent to which the effectiveness of the time-out technique is enhanced when combined with other behavioral treatments. (Hewat et al., 2006). Results from a new intervention approach for school-aged CWS (described in detail in Chapter 1.4.2) showed an average of 72% decreased in stuttering from the initial baseline to later stages. Additionally, the quantity of speech increased by 18%, naturalness scores improved by 50%, and the severity of stuttering decreased by 56%. Despite only four participants finishing the intervention, the benefits included both overall greater well-being and reduced disfluencies (Einarsdóttir et al., 2025).

The Lidcombe Program consists of two stages. Stage 1 is the *Therapy Stage*. In this initial stage, the focus is on providing immediate verbal contingencies for stutter-free speech. The child and the clinician engage in structured conversations or play activities, during which the clinician provides positive feedback (verbal praise) for fluent speech and gentle corrections for stuttered speech. The corrections are typically in the form of requests for the child "Can you say that smoothly". The aim is to increase the frequency of fluent speech and decrease the frequency of stuttering. Stage 2 is the *Maintenance Stage*. Once the child achieves a high level of fluency in Stage 1, they transition to Stage 2, where the focus shifts to maintaining fluent speech. The frequency of clinic visits decreases, and the child and clinician continue to monitor speech fluency and provide feedback as needed. The child's caregivers are also trained to implement the program at home, reinforcing the principles learned during therapy sessions. Intervention intensity and duration vary depending on factors such as the severity of stuttering, the child's age, and their response to therapy (Johnson et al., 2024; Jones et al., 2005; O'Brian et al., 2014; Onslow & Kelly, 2020; Onslow & Packman, 1999). Three recent systematic reviews on stuttering interventions for children and adolescents have identified the Lidcombe Program as having the highest level of evidence for treating young children who stutter (Brignell et al., 2021; Laiho et al., 2022; Sjøstrand et al., 2021). The effectiveness of the Lidcombe Program lies in its simplicity and parent-led nature, making it accessible for families to implement at home. Research has shown that the program is effective in various settings, including face-to-face sessions and telehealth applications (O’Brian et al., 2014). While the Lidcombe Program has primarily focused on preschool aged children, a recent study by Johnson et al. (2024) implemented a Phase II trial with 37 school-aged children, conducted via video telehealth. The results suggest that the program has the potential to benefit approximately one-third of children aged 6 to 12 years. The findings of this study are promising for school-age children. However, the pre-to-post-treatment design might have inflated the effect sizes. Without a no-treatment control group, it is unclear if the changes in outcome variables are directly due to the treatment. Furthermore, the study design does not consider the potential impact of other factors, such as the natural progression of the disorder (Johnson et al., 2024).

1.4.2 A stuttering intervention for older children

Stuttering Treatment for Older Children (STOC) is a new direct intervention approach for school-aged CWS. STOC is built on the work of Roger Ingham and his colleagues at the University of California, Santa Barbara, in 2011 (Roger Ingham, personal communication). The STOC intervention is built on the Response Contingent Time-Out framework and ideology (Hewat et al., 2006; James et al., 1989; Smith & Weber, 2016). The manual for STOC is available in the Supplementary Online Material of Einarsdóttir et al. (2024). The intervention consists of two daily 10-minute conversations between the parent and the child: one where the parent administered the intervention with parent-imposed time-out, and the other where the child practiced self-imposed time-out. In the parent-imposed time-out part of the STOC intervention the verbal contingencies are administrated by the parent to the child following a stuttering event. There is a small time-out period before the child starts speaking again. When the child is in the self-imposed time-out condition, the child applies the time-out following a stuttering event. There is no formal verbal contingency for stutter-free speech involved in the intervention, such as “good job”. The frequency of the daily sessions changes in accordance with stutter-free sessions, with adjustments based on the child’s fluency progress. When the child met the defined criteria for fluent speech (stuttering less than 1%SS, scoring 1–3 on the naturalness scale, and scoring 1–2 on the severity scale for three consecutive days) the training frequency decreased; otherwise, it was increased to the previous step in the intervention schedule. (see Section 3.2.4 for detailed description). With STOC, individuals who stutter need not alter their speech pattern, as is required in interventions focusing on fluency shaping, such as slowing down the rate of speech. Instead, emphasis is placed on awareness of stuttering events, with the administration of time-out contingent upon the presence of a stuttered event. Awareness training, in various forms, has been integral to stuttering interventions for decades, tracing back to Van Riper (1973), who utilized stuttering modification techniques such as pull-outs, where the speaker stops in the middle of a stuttered word and then smoothly finishes it, and cancellations, where the speaker pauses after a stuttered word and then repeats it more fluently, to replace instances of stuttering with controlled fluency.

The comparison between the STOC and Lidcombe Program interventions is important as it highlights the distinct mechanisms through which each approach addresses stuttering, despite their shared goal of reducing stuttering behaviour. While both interventions employ immediate consequences for stuttering occurrences, they do this in different ways. STOC consistently applies the time-out period after each stuttering event. In contrast, the Lidcombe Program encourages parents and clinicians to provide verbal contingencies for fluent speech more frequently than mild corrections for instances of stuttering (James et al., 1989; Johnson et al., 2024; Onslow & Packman, 1999). Both approaches aim to diminish stuttering behaviour by administering negative consequences for stuttering events, thereby reinforcing fluent speech (James et al., 1989; Johnson et al., 2024; Onslow & Packman, 1999). Moreover, both interventions can be tailored to meet the individual needs of CWS. Therapists or parents have the flexibility to adjust multiple parameters of the intervention delivery, such as the session length, session frequency, and amount of feedback, based on the child's progress and specific requirements (James et al., 1989; Johnson et al., 2024; Onslow & Packman, 1999). Furthermore, both STOC and the Lidcombe Program can be integrated into comprehensive stuttering therapy plans, allowing for combination with other intervention techniques such as speech restructuring, fluency shaping, and cognitive-behavioural strategies to address various aspects of stuttering (Onslow & Packman, 1999; Smith & Weber, 2016).

Table 1. Comparison of STOC and the Lidcombe Program

Aspect	STOC (Stuttering Time-Out Conditioning)	Lidcombe Program
Mechanism	Time-out from speaking initiated by parent or child after stuttering	Verbal contingencies delivered by clinician or parent
Application	Consistently applies time-out after each stuttering event	Provides verbal contingencies for fluent speech and mild corrections for stuttering
Goal	Diminish stuttering behavior by administering negative consequences for stuttering events	Reinforce fluent speech by providing positive feedback and mild corrections
Flexibility	Can be tailored to individual needs; intensity and frequency adjusted based on progress	Can be tailored to individual needs; intensity and frequency adjusted based on progress
Integration	Can be combined with other techniques like speech restructuring, fluency shaping, and cognitive-behavioral strategies	Can be combined with other techniques like speech restructuring, fluency shaping, and cognitive-behavioral strategies

1.5 Parent-implemented intervention for CWS

Given the profound impact of parent-child interactions on child development within daily activities, parents frequently play a crucial role in delivering interventions for various developmental disorders (Kaiser & Roberts, 2013; Valero-Aguayo et al., 2021; Wainer & Ingersoll, 2013). Research has shown that interventions involving parents, whether partially or fully carried out by them, often result in more favourable outcomes for children with disabilities, including those with autism spectrum disorder and intellectual disabilities (Ingersoll & Dvortcsak, 2006; Kaiser & Roberts, 2013; Valero-Aguayo et al., 2021). These parent-implemented interventions typically adhere to a *triadic intervention model,* which involves the practitioner, the parent, and the child (Druker et al., 2020; Lieberman-Betz, 2015; Salisbury & Cushing, 2013), wherein clinicians equip parents with tailored Intervention techniques, which the parents then use to help their children develop positive behavioural or skill changes (Lieberman-Betz, 2015). The success of these interventions depends not only on their inherent effectiveness but also on how well parents learn and apply the strategies, the regularity with which they use them, and the precision of their execution (Carroll et al., 2007; Ingersoll & Dvortcsak, 2006; Lieberman-Betz, 2015; Roberts & Kaiser, 2011; Salvo & Seery, 2021), i.e., their fidelity of implementing the intervention, or how well they are able to adhere to the intervention design.

In recent years, interventions carried out by parents for children who stutter have become increasingly prominent in academic research (Bergþórsdóttir & Ingham, 2016; Bothe et al., 2006; de Sonneville-Koedoot et al., 2015; Hofslundsengen et al., 2022). Parent-implemented interventions often encompass both direct and indirect approaches. Direct approaches, as exemplified by the Lidcombe Program (Johnson et al., 2024; Onslow et al., 2003), involve parents providing children with explicit instructions and feedback to enhance fluency. Conversely, indirect approaches such as the DCM (Franken et al., 2005), focuses on parents altering environmental demands, such as talking more slowly to the child, to reduce the likelihood of the child stuttering. Both the Lidcombe Program and DCM often utilize a triadic intervention model. In this approach, the SLPs provide intervention in clinical settings and training parents to deliver the intervention at home between appointments (Euler et al., 2021; Millard et al., 2008; Preston et al., 2022). Recent research has investigated various treatment methods for speech disorders, with mixed outcomes. Euler et al. (2021) studied speech restructuring techniques with 119 participants, finding that 16.8% experienced no noticeable treatment benefit. Johnson et al. (2024) evaluated operant programs in a study involving 37 participants, revealing that 67.7% showed a partial response, while 13.5% showed no response. Kohmäscher et al. (2023) examined stuttering modification methods, noting significant variability in fluency outcomes and minimal long-term effects. Although researchers, clinicians, and individuals who stutter continue to debate the effectiveness and suitability of various intervention approaches (Nippold, 2012; Yaruss et al., 2012), FOI considerations are often overlooked (Brignell et al., 2021; Hofslundsengen et al., 2022).

1.6 The importance of ensuring FOI in research for CWS

While various therapeutic approaches have demonstrated efficacy for CWS (e.g., Andrews et al., 2016; de Sonneville-Koedoot et al., 2015), there exist numerous factors contributing to reductions in stuttering among CWS that may be independent of intervention effects. The variability in recovery rates documented across studies underscores the complexity of determining intervention effectiveness, particularly considering the high rate of spontaneous recovery among CWS (Blood & Blood, 2004; Einarsdóttir et al., 2020; 2024; Yairi & Ambrose, 2005). Spontaneous recovery, defined as the natural decline in stuttering frequency or severity over time without formal intervention, can confound intervention outcomes by either masking the effectiveness of non-effective interventions or attenuating the apparent impact of effective ones (Bergþórsdóttir & Ingham, 2016; Ingham & Riley, 1998). Additionally, individual characteristics of CWS, including age of onset, severity of stuttering, and comorbidities such as language disorders, further complicate the evaluation of intervention effectiveness (Yairi & Ambrose, 2013). Without consistent and accurate application of the intervention, it becomes difficult to distinguish whether the improvements are due to the treatment itself or other external factors. Additionally, the assessment of study design is crucial, as differences in doses between arms or the lack of a control group to account for natural recovery can further complicate the interpretation of results. These factors highlight the importance of rigorous and well-controlled studies, making sure that the observed outcomes are genuinely attributable to the intervention and not influenced by external variables. Maintaining and monitoring FOI throughout the intervention helps attribute outcomes to the intervention itself, rather than confounding variables.

The complex nature of stuttering and its interventions mean that clinicians and researchers encounter significant challenges in confidently interpreting intervention outcomes. Therefore clear and explicit measurements of FOI are essential. Measurement of FOI stands as a pivotal aspect of studies seeking to evaluate the effectiveness of such interventions, as it encompasses not only the efficacy of the intervention but also the delivery process (Bellg et al., 2004; Carroll et al., 2007; Hofslundsengen et al., 2022). FOI refers to the extent to which an intervention is delivered as intended and is a critical determinant of intervention effectiveness (Carroll et al., 2007; Dane & Schneider, 1998; Dusenbury et al., 2003; Lieberman-Betz, 2015; Perepletchikova et al., 2007). By meticulously assessing FOI, researchers can enhance outcome comparability across studies and bolster confidence in the reliability of study findings (Perepletchikova et al., 2007; Moncher & Prinz, 1991). While effective interventions hold promise for improving outcomes in CWS, the inherent complexities of stuttering and its intervention necessitate careful consideration of various contextual, individual, and methodological factors. Moreover, transparent reporting of FOI measures allows for critical appraisal of study methodology and facilitates replication of effective interventions in clinical practice (Hofslundsengen et al., 2022; Schoenwald et al., 2011). Clear and explicit measurement of FOI serves as a crucial safeguard in stuttering intervention research, ensuring rigor and reliability in the evaluation of intervention effectiveness.

As interventions with a parent-implemented component are often administered across multiple settings; clinics, educational settings, and home settings, it is both challenging and imperative to document FOI across all delivery contexts and by all intervention agents (An et al., 2020; Hofslundsengen et al., 2022). A comprehensive consideration of FOI in all intervention contexts fosters standardization in implementation, thereby rendering interventions more conducive to research and replication (Mowbray et al., 2003). Moreover, systematic measurement of FOI allows researchers to identify areas of intervention delivery that may require improvement or refinement. By examining fidelity across different settings and intervention agents, researchers can gain insights into the factors influencing intervention implementation and adapt strategies accordingly. Additionally, documenting FOI promotes transparency and rigor in research practices, enhancing the credibility and reliability of study findings (Bergþórsdóttir et al., 2022). Furthermore, understanding the nuances of FOI in diverse contexts can inform the development of guidelines and best practices for intervention delivery. By elucidating the components and requirements for competent implementation across different settings, researchers can establish benchmarks for fidelity assessment and promote consistency in intervention delivery practices (Hagermoser Sanetti & Kratochwill, 2014). This, in turn, facilitates comparability across studies and enhances the cumulative knowledge base in the field. While some differences between intended and achieved fidelity are expected, the extent of these differences is crucial, as it can often explain why an intervention was unsuccessful or why its effects varied across participants (Bellg et al., 2004; Hagermoser Sanetti & Kratochwill, 2014). In summary, comprehensive measurement of FOI is essential for evaluating the effectiveness of interventions, promoting standardization in implementation practices, and advancing research in the field. By systematically assessing fidelity across various contexts and intervention agents, researchers can optimize intervention delivery, enhance research validity, and ultimately improve outcomes for individuals receiving interventions.

1.7 FOI measurements in research for CWS

To date, systematic collection and reporting of FOI measurements in interventions for CWS has been lacking (Bergþórsdóttir et al., 2022; Hofslundsengen et al., 2022), and these measurements have not been explicitly utilized to support the interpretation of intervention outcomes. A recent review of 36 intervention studies published between 1990 and 2020, incorporating a parent-implemented component for CWS, alongside clinical sessions, examined the presentation of FOI information in these studies, categorizing reporting as evidence/direct measurement, based on a report, unclear, or not reported (Bergþórsdóttir et al., 2022). Regarding whether studies provided evidence of fidelity for parent-implemented sessions, evidence was only available in 19.4% of studies for dosage, 5.6% for adherence, 8.3% for quality, and responsiveness was never reported. Across studies the component most reported on was dosage, and responsiveness was least reported on. Notably, none of the 36 studies examined or reported on all four FOI components for both clinical and home interventions The quality of FOI reporting did not increase over time. These findings are in line with the results found in the systematic review done by Hofslundsengen et al. (2022). The review included 21 stuttering intervention studies, published between 1974 to 2019. All studies reported on details regarding the description of the intervention program, either by clear description or referral to manual (Hofslundsengen et al., 2022). Hofslundsengen et al. (2022) reported that 50% of the studies provided unclear or no details of dosage in home settings and for adherence only in the home settings, 83% were reported to have unclear or no details provided. Borrelli et al.'s (2005) extensive evaluation across 10 years of Health Behavior Research, found that out of the 342 included articles, 22% detailed strategies for maintaining provider skills, 27% included checks for adherence to protocol, and 35% utilized a treatment manual. Notably, 54% of the articles did not report using any of these strategies, while 12% employed all three. On average, the adherence to treatment fidelity strategies was 0.55, with 15% of the articles achieving an adherence rate of 0.80 or higher. While our study found that the use of these strategies did not significantly change over time, Borrelli (2005) reported similar findings, indicating a consistent trend across different research contexts.

Recent systematic reviews on intervention studies for people who stutter have commented on FOI in the studies included in their reviews. Nye and Hahs-Vaughn (2011) reviewed 23 studies published between 1969 to 2008 examining the methodological quality of stuttering intervention studies for children and adults. They found that FOI was poorly reported on, with only half of the studies including adequately detailed descriptions of interventions, 35%

of the studies did not explicitly describe the intervention setting or context, limiting the replicability and generalizability of the findings. Baxter et al. (2016) evaluated 137 studies published between 1990 and 2020, focusing on the clinical effectiveness of interventions for people who stutter. They conclude that to gain better clarity on the benefits of the interventions, there needs to be an enhanced understanding and measurement of FOI in the intervention process, stating that by accurately measuring FOI, researchers and clinicians can pinpoint the elements that are most effective, leading to more consistent and successful intervention results. Finally, Brignell et al. (2021) reviewed 20 studies concerning direct and indirect interventions for children and adolescents who stutter. They stated that future research needs to consider FOI, the reporting of which was found to be lacking in the current literature. These conclusion about FOI from systematic reviews of intervention studies underscore the paramount importance of precise FOI measurements to maintain the consistent efficacy and reliability of interventions for stuttering, for both children and adults.

These findings underscore the challenge of establishing whether interventions contributed to recovery in the absence of comprehensive knowledge regarding their implementation, including potential factors such as spontaneous recovery. Additionally, given that interventions for CWS are predominantly conducted by parents in home-based sessions, the absence of intervention delivery by trained professionals, such as SLPs, underscores the critical importance of thorough fidelity measurement. Measuring only dosage may overlook differences in adherence, quality, and responsiveness resulting from variations in parental delivery compared to professional delivery. Furthermore, the variability in parental training, understanding, and consistency in applying the intervention techniques can lead to significant differences in outcomes. This highlights the need for developing robust training programs for parents and creating mechanisms for ongoing support and monitoring to uphold fidelity. Without such measures, it is challenging to attribute improvements in stuttering directly to the intervention, as opposed to natural variability or external factors.

2 Aims

The overarching aim of this thesis was to evaluate FOI components in the context of stuttering interventions and evaluate the relationship of FOI components to outcome measurements of stuttering interventions.

The aims of this thesis were to:

1. Gain a comprehensive understanding of the reporting of FOI measurements in previous stuttering intervention research for CWS where a component of the intervention was implemented in a home environment and to identify gaps in knowledge related to FOI measurement and reporting (Paper I & II, Study I).

2. Examine the effectiveness of a pilot home-based intervention for school-aged CWS (stuttering treatment for older children [STOC]) with detailed measurements of fidelity of implementation (FOI) incorporated into the research design. (Paper III, Study II).

3. Assess whether high levels of fidelity can be achieved in an intervention study for school-aged CWS where FOI was incorporated into the design and tightly controlled for during the study (Paper IV, Study III).

4. Determine the relationship between measured FOI and performance on outcome measures in an intervention study for stuttering in school-aged CWS (Paper IV, Study III).

2.1 Study I: FOI in parent-implemented interventions for stuttering

Study I aimed to review the reporting of the four key FOI components (dose, adherence, quality, responsiveness) in published intervention studies of parent-implemented intervention for school-aged CWS. The motivation for this study came from the lack of prior focus of fidelity in interventions used with CWS and the possibility of wrongfully assessing an intervention as successful or unsuccessful (Paper I). The following research questions were addressed in a review following a systematic literature search (Paper II): (a) How frequently are the four FOI components reported in studies where interventions are administered by clinicians versus parents? and (b) Has the reporting of FOI components evolved over time? This study aimed not to assess the efficacy of the interventions reviewed, but to examine the extent and quality of FOI measurement and documentation in previous research.

2.2 Study II: A pilot study of stuttering treatment for older children (STOC)

The aim of the second study in this thesis (Paper III) was to establish how well participants responded to a new intervention for school-aged CWS, STOC. The aim was to examine individual responsiveness to STOC, using a single case experimental design, and to investigate changes in stuttering-related quality of life. Fewer participants than expected were encountered, as it was challenging to find many individuals in this age group in Iceland who stutter (3% SS) within a clinical setting. Given that a new approach is being piloted, it was important to closely monitor the changes for each participant. Therefore, employing a multiple baseline design was deemed an appropriate methodology. The following research questions were addressed: (a) Does participation in STOC decrease the frequency of stuttering of school-aged CWS during and after intervention. (b) Does participation in STOC result in changes in quality of life?

2.3 Study III: The effect of FOI in a pilot study of stuttering treatment for older children

Study III examined the relationship between four FOI measures and children's outcomes using data from the STOC intervention study for CWS (Paper IV). This study addressed the following research questions: (a) What is the relationship between each FOI measurement (dosage, adherence, quality, and responsiveness) and each outcome measurement (%SS, Syllables Per Minute (SPM), naturalness, and severity) for children participating in the STOC pilot study? (b) Do differences exist in the FOI measurements between participants who completed the STOC intervention program and those who did not? (c) What level of fidelity was achieved in the STOC study?

3 Materials, Methods, and Results

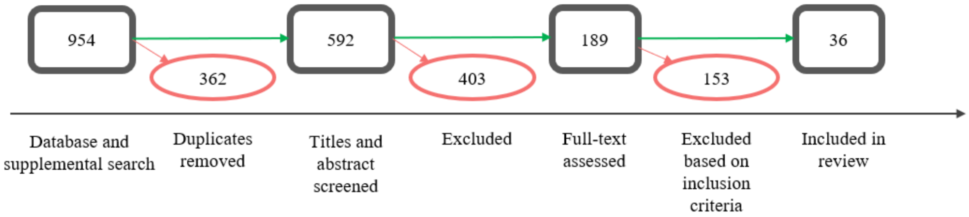
3.1 Literature review (Study I)

3.1.1 Data sources

In February 2021, a thorough search was conducted across several electronic databases including CINAHL, PubMed, Scopus, and Web of Science. The aim was to gather evidence regarding implementation fidelity in published reports of joint clinician- parent implemented interventions for CWS, across both clinical and home settings. In total 945 records were found through this search, of which 592 were unique. In addition to the database search, a supplemental hand search was conducted of reference lists of included articles and expert advice. Following this, expert advice was sought as to whether any relevant articles had been missed. Nine additional relevant articles were found through supplemental searching.

Selection criteria: The first screening involved reviewing the titles and abstracts of all the records using the following inclusion criteria. Studies had to (a) be published in English, (b) be published in peer reviewed journals, (c) be published after 1981, (d) describe empirical treatment studies of behavioural interventions for CWS, (e) include a parent-implemented intervention with CWS, and (f) have participants who were CWS aged 6 years or younger. The year 1981 was selected as a cutoff because it marks the publication of Yeaton and Sechrest's (1981) outline of the critical dimensions of treatment fidelity. This publication is considered a pivotal point in the study of fidelity in intervention research. Title and abstract screening excluded 403 records. A full text screening based on the inclusion criteria was completed for the remaining 189 recordings leaving 36 articles in the review. The screening process can be seen in Figure 1.

Figure 1. The Process of Data Selection



Data extraction and coding: Key information was extracted from each article by the author of this thesis, including publication year, participant characteristics (sample size, age), methodology (design), and intervention details (name/type) (see Appendix C). Each article was coded to determine the presence or absence of four FOI components for both clinician-implemented and parent-implemented parts of interventions: (a) dosage, (b) adherence, (c) quality, and (d) responsiveness. The coding scheme was developed by the first and second authors based on review of the literature on FOI measurement, see Appendix in Paper II for these scales. This defined each FOI component (see Chapters 1.1.1 – 1.1.4).

The presence of each FOI component was based on pre-specific criteria. For example, dosage was considered present if the number of sessions attended, session length, or intervention duration was reported. Adherence was marked present if any measurement of accuracy in intervention delivery was mentioned, such as the use of logbooks. Quality was noted as present if the article discussed the level of delivery of the intervention strategies, for example if the researchers observed this in recordings of intervention sessions. Responsiveness was assessed as present if any reference was made to the reactions interventionists (clinicians/parents) or participants (children) in intervention sessions, including their comprehension and engagement. Coding was binary (present/absent), with any mention of an FOI component considered as present even if data were not reported (see Appendix C). The strength of evidence for each FOI component was coded as none, unclear, based on report, or evidence/measurement, indicating the level of detail provided.

3.1.2 Reliability

Selection: The thesis author reviewed the titles and abstracts of identified articles for suitability for this review, adhering to the specified inclusion criteria. Out of the total 592 records, 403 were excluded while 189 met the criteria. Subsequently, two hundred of the total records (33.3%) were randomly selected and their titles and abstracts were re-screened by a committee member (KC). There was 100.0% agreement between the final status (excluded or included) of each record.

Full text screening: Applying the aforementioned inclusion criteria, the author of this thesis and her supervisor (JTE) independently both reviewed the full texts of the remaining 189 articles. The inter-rater reliability was 94.0% (*n*=178), with discrepancies on the inclusion of the remaining 6.0% (*n*=11) papers resolved through discussion until a consensus was achieved.

Coding of data: FOI variables were coded as either present or absent. If an article indicated that a particular aspect of FOI had been assessed, it was categorized as present, irrespective of whether the data for that FOI component was explicitly reported in the article. The level of evidence supporting each FOI component was categorized into four levels (none, unclear, based on report, evidence/direct measurement) for both clinician-implemented and parent-implemented interventions. These categories were defined as follows: *None* indicated no evidence or discussion of the FOI component in the article; *Unclear* suggested mention of the FOI component, but without details on the method or outcome of examination; *Based on report* indicated the authors' claim of evaluation without reported data; *Evidence/direct measurement* represented data describing the FOI component in the intervention delivery reported in the article. All articles (100.0%) were independently coded by the thesis author and her supervisor (JTE). Point-by-point reliability was 95.3%, with discrepancies resolved through discussion to reach consensus. Borrelli et al. (2005) employed a similar coding scheme, except they included an additional category 'not applicable'. This category was used when a particular treatment fidelity strategy was not relevant to the article in question. However, this category was not considered necessary in the current study, as all interventions should consider the FOI components.

3.1.3 Results of FOI measurements in literature review

The 36 studies spanned the years 1990 to 2020 (*M*=2011, *SD*=7.5) and described a total of 926 CWS. Participant numbers varied across studies, ranging from one to 199 children (*M*=15.5, *SD*=36.9). The age range of participants were reported in 33 studies, with the reported range from 2;0 to 6;3 years. Among the remaining three studies, one indicated that children were aged ≤5;11, while two reported mean participant ages as 4;2-4;3 years and 3.7 years, respectively. A total of nine different interventions were investigated, with some studies exploring modifications to standard intervention protocols (e.g., delivery via telehealth). Four studies examined multiple interventions. Research designs employed in the studies comprised Experimental Group Comparisons (*n*=19, 52.8%) or Single-Case Experimental Designs (*n*=17, 47.2%).

The reported components of FOI for interventions conducted in clinical settings and in home sessions can be seen in Table 2. As shown in the table, dosage is the most frequently reported component in the clinical setting, whereas adherence and quality are the least reported. Conversely, in the home setting, adherence is the most commonly reported component, while responsiveness is the least reported component.

Table 2. Reported FOI Components of the 36 Studies

	FOI components reported *n* (%)			
FOI component	Dosage	Adherence	Quality	Responsiveness
Clinical	26 (72.2%)	7 (19.4%)	7 (19.4%)	11 (30.6%)
Home	7 (19.4%)	13 (36.1%)	11 (30.6%)	7 (19.4%)

When looking at each study to evaluate how many of the four components each study is reporting on, the number of FOI components ranged from 0 to 4 (*M*=1.42, *SD*=1.18) in clinical settings and 0 to 4 (*M*=1.03, *SD*=1.18) in home settings (see Table 3). In home settings most studies were evaluated to not report on any of the components (*n*=17, 47.2%), while this was most often one FOI component reported in clinical settings (*n*=14, 38.9%). The results show that three studies report on all four FOI components in the clinical setting and two studies report on all four FOI components in the home setting. For the clinical setting 61.1% report on zero or one FOI component and in the home setting this is 66.1%.

Table 3. Number of FOI Components Reported on in Clinical and Home Settings (*N*=36)

	FOI components reported *n* (%)				
Setting	Zero	One	Two	Three	Four
Clinical	8 (22.2%)	14 (38.9%)	8 (22.2%)	3 (8.3%)	3 (8.3%)
Home	17 (47.2%)	7 (19.4%)	7 (19.4%)	3 (8.3%)	2 (5.6%)

When looking at the combined reports of FOI in both clinical and home settings together, 10 (27.8%) studies reported on half or more than half of the necessary FOI measures, and no studies reported on all eight components, based on that all the studies were evaluating both clinical and home settings, leaving each study with eight FOI components to be evaluated. In total for the 36 studies when looking at clinical and home settings together, there were reports of the following FOI components: dosage (*n*=33, 45.8%), adherence (*n*=20, 27.8%), quality (*n*=18, 25.0%), and responsiveness (*n*=18, 25.0%). Five of the studies did not report on any FOI components.

When looking at the reporting strength of FOI components in each study (none, unclear, based on report, evidence/direct measurement), the majority of articles were categorized as “not reported” for both clinical and home setting for all FOI components as can be seen in Figures 2 and 3. The only exception was for dosage in clinical settings where 26 (72.2%) articles were classified as having evidence/direct measurement. When present, information on adherence was based on report (clinic *n*=7, 19.4%; home *n*=8, 22.2%). Similarly, information on quality, when present, was most often based on report (clinic *n*=5, 13.9%; home *n*=6, 16.7%). Finally, when information on responsiveness was present it was also based on report (clinic *n*=7, 19.4%; home *n*=4, 11.1).

Figure 2. Reporting Strength of FOI in Clinical Settings

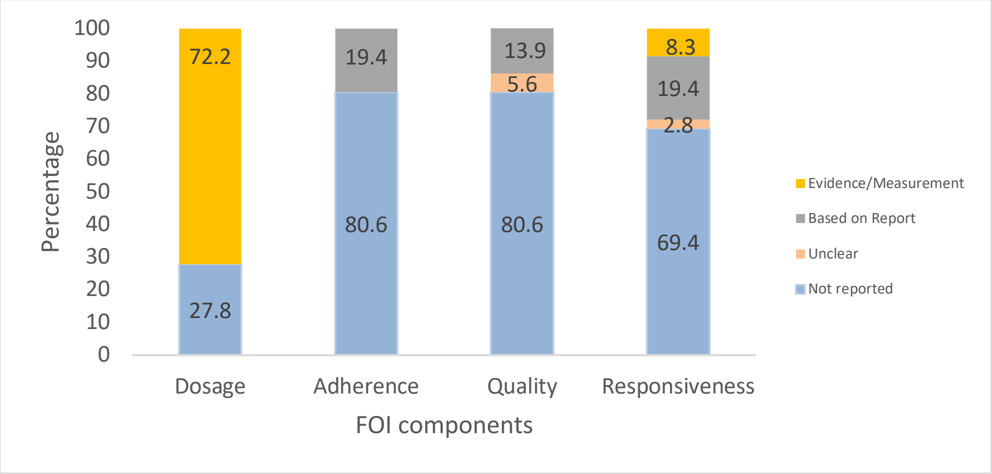
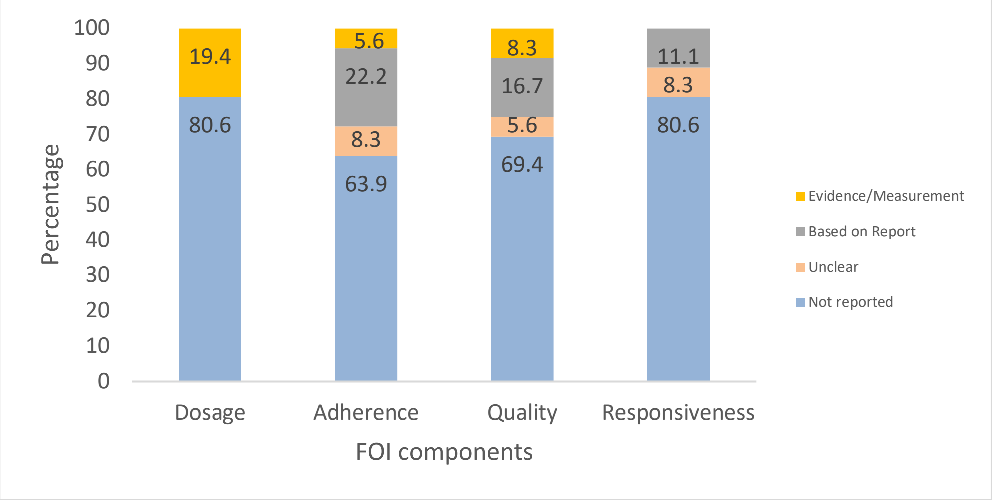


Figure 3. Reporting Strength of FOI in Home Settings



A Spearman rank-order correlation was performed to examine the relationship between the year of article publication and the number of FOI components reported. Each of the four components was coded as present or absent for both clinical and home settings, resulting in a possible maximum score of eight. The analysis revealed no significant relationship between publication date and the number of FOI components reported, r_s (36) = .044, p = .80.

3.2 Pilot STOC study and effect of FOI (Study II and III)

3.2.1 Ethical approval

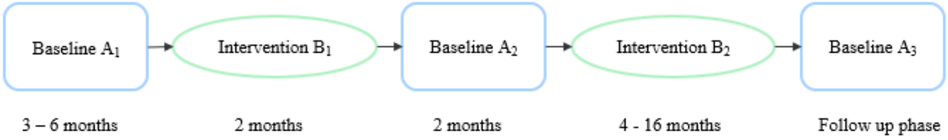
Ethical approval was obtained from the Icelandic Bioethical Committee 2014 registration number (S14-059).

3.2.2 Experimental design

This study used a single case experimental design, using a withdrawal/reversal design with an added randomised multiple-baseline design, as well as a prolonged follow-up (see Figure 4). A withdrawal/reversal design means that participants were able to serve as their own controls, using within-participant comparisons of data gathered in baseline and intervention phases. A risk with using this design is that the highly variable nature of stuttering could reduce the chance of a clear relationship between the independent variable (STOC) and the dependent variables (outcome measures) emerging. To counter this, additional elements were added to the experimental design for extra security, so that if a relationship between the independent and dependent variables existed, it would be able to be seem. First, given the highly variable nature of the occurrence and severity of an individual's stuttering over time, i.e., temporal variability (Constantino et al., 2016), randomised multiple baselines were used. Randomised multiple baselines used case randomisation and intervention start-point randomization so that participants had baseline periods of different lengths, meaning they started intervention at different points (Levin & Ferron, 2021). The multiple baseline measurements consisted of five different contexts in which data for each of the four dependent variables were collected to account for the highly variable nature of the occurrence and severity of in individual's stuttering across different speaking situations i.e., situational variation (Guttormsen et al. 2021). This design feature minimised the risk that (a) observed changes in the dependent variables were the result of increased familiarity, and resulting decreased stuttering, with the speaking tasks used to measure the dependent variables, and (b) any one dependent variable measure was a context in which a child habitually had an unrepresentatively low or high occurrence of stuttering. Finally, a prolonged follow-up period was included as a major criticism of period intervention studies for stuttering has been that these studies only include short term and immediate post-intervention outcome measurements (Johnson et al., 2023; Sjøstrand et al., 2021). The elements included in the single case experimental design used in this study exceed the elements that are deemed necessary for a quality single case experimental design (Council for Exceptional Children, 2014; Ganz & Ayres, 2018; Tate et al., 2016), and in some regards exceed what is considered to be *aspirational* for recommended methodological standards (Ganz & Ayres, 2018).

Participants were randomly assigned to varying lengths of baseline A_1: three months (P2, P3, P7), five months (P1, P5, P6), or six months (P4), i.e., randomised multiple-baseline design. After the baseline phase, all participants entered a two-month intervention period (B_1). This was followed by another two-month baseline period (A_2). Then, participants underwent a second intervention block (B_2), with the duration varying for each participant based on their response to the intervention depending on how well they were managing their stuttering i.e., withdrawal/reversal design. Finally, the follow-up phase (A_3) involved collecting samples over the nine months after the intervention ended, with monthly samples collected for three consecutive months and one additional sample collected six months later, i.e., prolonged follow-up design.

Figure 4. Single Case Experimental Design: Baseline and Intervention Phases



3.2.3 Participants

Seven monolingual Icelandic-speaking boys aged 9 to 13 years (*M*=11;5, *SD*=1;7) were recruited for the study based on referrals from SLPs or special education teachers. The 9-13 age group was chosen to minimize the confounding effects of spontaneous recovery, which is more common in preschool children, allowing for a clearer assessment of the intervention's efficacy. Although female participants were referred, none met the inclusion criteria. Inclusion criteria for participation in the STOC pilot study included: (a) producing at least 3% of syllables stuttered (%SS) in clinical settings during oral reading or conversational speaking tasks; (b) expressed concern about the child's stuttering from both the parents and the child; and (c) willingness of both parents and child to take part in the study. Exclusion criteria were having a severe developmental language disorder or developmental delay. Parent-reported onset of stuttering ranged from 2 to 10 years of age (*M*=4;5, *SD*=2;8). The range of age at the start of intervention was from 9;5 years to 13;9 years (*M*=11;6, *SD*=1;7). Range of %SS at baseline was from 2 to 17.7 (*M*=8, *SD*=6.8), range of SPM at baseline was 82.6 to 192.6 (*M*=138, *SD*=40). For naturalness at baseline the range was from 3 to 9 (*M*=5.4, *SD*=2.2), (1 = highly natural; 9 = highly unnatural) and for severity at baseline the range was from 3 to 9 (*M*=5.5, *SD*=2.1), (0 = no stuttering, 10 = severe stuttering) as defined in Section 3.2.6.

3.2.4 Procedure of the STOC intervention

The STOC intervention, as outlined in the manual (available in the Supplementary Online Material to Einarsdóttir et al. (2024), consisted of two 10-minute conversation sessions per day, one in the morning utilizing parent-imposed time-out and another in the afternoon/evening utilizing self-imposed time-out. Whenever a stuttered event occurred, either the parent verbally prompted the child to pause by saying “stop” or the child spontaneously paused from speaking for 2-3 seconds before continuing.

Intervention training: Prior to the first intervention period (B_1), participants and their parents underwent training on implementing time-out in a clinical setting. Individualized training materials were developed based on edited video recordings of each participant's baseline recordings. Initially, participants and parents were trained to recognize stuttering events in the recordings. Subsequently, during conversational practice sessions, parents practiced prompting the child to pause when a stuttered event occurred, followed by the child practicing self-interruption when they identified that they had stuttered. The duration of the conversational training phase of the intervention training ranged from 1 to 4 sessions, depending on the time required for parents and children to accurately identify and respond to stuttered events in at least 80% of occurrences. Training sessions were usually an hour in duration, including time for feedback and discussion. The 80% threshold was picked because achieving 100% accuracy in identifying and responding to stuttered events during conversation is impractical. This threshold is deemed sufficient under the training conditions, despite the lack of specific rationale in the literature, such as in the Lidcombe Program, which focuses on teaching verbal contingencies rather than accuracy in identifying stuttered events (Onslow et al., 2024). In cases where the adherence of parents and/or children dropped below 80% for four consecutive sessions during intervention periods (B_1, B_2), they underwent retraining sessions (one hour in duration) until their adherence was above 80%. Retraining sessions were conducted either in person or via Skype.

Judge training: Seven judges assessed participants' stuttering throughout the data collection and analysis phases. Among these judges were two certified SLP’s and five graduate students in speech-language pathology serving as research assistants. All judges received 15-20 hours of training in stuttering assessment using the Stuttering Measurement System (SMS) program (Ingham & Ingham, 2011), the severity scale (see Appendix to paper III) and using Icelandic samples available on the STUREN homepage (https://sturen.vercel.app/; Einarsdóttir et al., 2014).

Intervention implementation: When the child met the defined criteria for fluent speech during intervention sessions, the training frequency decreased. The criteria for fluency were: (a) stuttering less than 1%SS, (b) scoring 1–3 on the naturalness scale, and (c) scoring 1–2 on the severity scale. These conditions had to be met for three consecutive days (i.e., across six consecutive training sessions). If a participant did not meet the defined criteria for fluency, the session frequency was increased to the previous step in the intervention schedule (see Supplementary Material Paper III). Trained judges reviewed each intervention session recording, as recorded through the app (see Section 3.2.6), and provided feedback via email to the parents within 24 hours. Feedback focused on stuttering frequency and adherence to the Protocol for Clinician-Managed Stuttering outlined in the STOC manual. The intervention frequency was adjusted based on the participant's performance during intervention sessions (B_1 and B_2). If the criteria for fluent speech were met (stuttering less than 1%SS, naturalness score of 1–3, severity score of 1–2) for three consecutive days (i.e., over six consecutive training sessions), then the training frequency decreased. The changes in session frequency when the criteria for fluent speech were met self-imposed time-out and parent-imposed time-out sessions were incrementally decreased to every second, third, fourth, fifth, sixth then seventh day. If the fluency was still maintained when sessions were every 7th day, then the session frequency was decreased to every 14 days and then monthly. Conversely, if the criteria were not met, the session frequency increased using the same increments to a maximum of once daily parent-imposed time-out and self-imposed time-out sessions. Researchers informed parents of any changes to the frequency of intervention session via email. Participants could withdraw from the intervention at any time by directly communicating with the researcher. Upon withdrawal, they were evaluated as "not completed".

3.2.5 Fidelity of implementation components

FOI was assessed across four parameters for sessions where the parent prompted the child to pause from speaking after a stuttering event occurred (parent-implemented STOC session) and sessions where the child self-imposed a pause from speaking following a stuttering event (child-implemented STOC session). The FOI elements examined were dosage, adherence, quality, and responsiveness, following specific definitions and procedures. Dosage was determined by comparing the proportion and length of planned sessions to those delivered, with records kept of each scheduled intervention session and every session recorded on video. Adherence, the accurate implementation of the intervention, was evaluated based on the accuracy with which the parent and the child identified and responded to stuttering events during sessions. Quality was based on how parents managed communication in the sessions to allow for children to produce more speech and thus have more opportunities to practice the intervention technique, e.g., the parent asking open questions rather than yes/no questions. Judges assessed the flow of conversation in each session using a four-point scale. Ratings were made by trained judges who reviewed each 10-minute session video, see Appendix in Paper IV. Responsiveness was coded by rating the child's happiness and engagement during sessions on a three-point scale, with assessments conducted by trained judges, see Appendix in Paper IV. These scales have not been validated. While in some studies the accuracy of intervention implementation is considered as a part of quality (i.e., the accuracy with which parents/children imposed a time out after a stuttering event), this was not considered to be a component of quality or responsiveness in this study, as this is specifically examined under the component of adherence. All recordings were scored for FOI according to the STOC study protocol, except when the recording schedule changed, in which case recordings were scored accordingly. Based on the findings from the literature review on commonly neglected areas in FOI investigations for treatment studies of children who stutter, this study examined the same FOI elements.

3.2.6 Outcome measures

To assess the intervention's impact and control for speech performance variability before, during, and after the intervention, monthly video recordings were gathered throughout all five phases of the study (A_1 - B_1 - A_2 - B_2 - A_3) (following the example of Ingham et al., 2012). The SMS program (Ingham et al., 2005) was utilized to compute outcome measurements, including %SS, SPM, naturalness, and severity. Data collection spanned from May 2014 to June 2017, resulting in 653 monthly recordings totalling nearly 80 hours. Calculations were done by the SMS program, %SS was calculated by dividing the number of stuttering occurrences by the number of spoken syllables, while SPM was determined by dividing the number of syllables spoken by three, representing the minutes of speech as three-minute samples of each recording was analysed. Naturalness was rated using a 9-point ordinal scale (1 = highly natural; 9 = highly unnatural) (Martin et al., 1984). Severity was rated on an 11-point ordinal scale (0 = no stuttering, 10 = severe stuttering) (Einarsdóttir et al., 2014).The trained judges described above did the ratings for naturalness and severity after listening to the recordings.

Speech fluency measures: Monthly fluency measurements were recordings of five distinct speaking scenarios: (a) reading aloud a passage tailored to the participant's reading level and interests, (b) delivering a three-minute monologue, (c) engaging in a ten-minute conversation with a friend or relative (d) participating in a ten-minute telephone conversation, and (e) engaging in a ten-minute speaking task that the participant chose themselves and perceived as challenging. Thus, a total of 36 minutes of speech recordings were collected monthly from each participant. Each recording was evaluated for the number of syllables, instances of stuttering, SPM, %SS, and severity (Einarsdóttir et al., 2014) and naturalness (Martin et al., 1984) ratings were made.

Intervention session measures: Measurements of the children’s speech in each session were made by evaluating videos recorded on iPod Touches equipped with a custom-designed app (see Paper III). Within the app the sessions were recorded and the app limited the duration of each treatment session to 10 minutes. It only allowed two sessions a day to be recorded (one in the morning and one in the evening). Recordings were delivered via the app directly to the researcher team. Then the parent and the child independently judged whether stuttering had occurred during the session and rated the naturalness of the speech in that session (Martin et al., 1984). Videos and this information from parents and children were sent directly to the researchers through the app. For every session the metrics number of syllables, instances of stuttering, SPM, percentage of syllables stuttered (%SS), severity, and naturalness, were evaluated based on the first three minutes of the child's speaking time in each session. A total of 1551 intervention recordings were received and assessed for the seven participants, with the number of recordings submitted by participants ranging from 148 to 350 recordings (*M*=222, *SD*=81), on average, 83.3% of the planned treatment sessions were delivered.

Quality of Life: Measurement of changes in quality of life before and after intervention involved the administration of two different questionnaires during the collection of the first and final speech samples. The first questionnaire utilized an Icelandic translation of the Speech Performance Questionnaire (SPQ) (see Boberg & Kully, 1994) adapted for this study with five questions focusing on participants' self-perceptions of fluency and speech. This questionnaire, previously used with adult and adolescent stutterers in Canada and the Netherlands (Langevin et al., 2006) and the USA (Ingham et al., 2015), was employed in our study to capture school-age children's perspectives on treatment-related changes in speech fluency and the effort involved in speaking, despite the lack of a comparable questionnaire for this age group. These questions included inquiries about current speech fluency, ability to sound normal when controlling speech, ease of speaking normally without conscious control, preference for stuttering over controlled speech, and self-identification as a stutterer. The second questionnaire employed an Icelandic translation of the OASES-S (Leósdóttir, 2014), tailored for children aged 7-12 years, which assesses experiences of stuttering and its impact on quality of life across four domains: general information, reactions to stuttering, communication in daily situations, and overall quality of life. The evaluation was done both pre- and post-intervention.

3.2.7 Reliability

Study II To ensure reliability, three SLP graduate students, uninvolved in the initial scoring, rescored 88 out of 653 (13.5%) monthly recordings and 116 out of 1551 (7.5%) intervention sessions. Systematic random sampling was utilized for the selection of monthly recordings, across participants and various speaking scenarios. Initially, a random number generator in Excel was employed to determine the first recording, selecting a number between 1 and 7. Subsequently, every 7th recording was chosen, resulting in an average of 13 recordings per participant (Range=7-24). This method was similarly applied to the intervention session recordings selected across participants and morning and evening sessions, with every 13th recording being selected. Yielding an average of 17 intervention recordings per participant (Range=10-23). High inter-rater reliability was observed through intraclass correlation coefficients: .92 for %SS, .87 for SPM, .78 for naturalness, and .78 for severity (Koo & Li, 2016).

Study III Dosage assessment involved comparing the actual number of intervention sessions with the expected number. Upon re-evaluation by the thesis author, both sets of values for each child were found to be accurate, resulting in a 100% match. For adherence, all 1471 (100%) recordings were independently rated by two judges, resolving discrepancies through discussion until consensus was reached. As the assessment of quality and responsiveness in the 1409 (95.8%) recordings was assessed by two judges (non-overlapping). The thesis author evaluated 161 (10.4%) of the recordings with an equal proportion of videos scored by each judge represented. Point-by-point reliability was 95.3%.

3.2.8 Data analysis

Study II All data for this study were compiled and organized in a Microsoft Excel spreadsheet for systematic analysis. Each session’s ratings were recorded in alignment with the video submission dates. In cases where a video recording was missing, no data entries were made for that session, and no adjustments or imputations were made for missing data in the analyses. This approach aimed to preserve the integrity of the data and reflect the actual availability of each session. Each rating was completed on the same day the video was received or, at the latest, within one week of submission. This protocol ensured that ratings were as accurate and contextually relevant as possible, minimizing recall bias by keeping the assessment period close to the video review date. All recordings were evaluated by following metrics: the number of syllables, the number of stuttering occurrences, SPM, %SS, severity scale rating (Einarsdóttir et al., 2014), and naturalness scale rating (Martin et al., 1984).

Study III All data analysis was conducted using Microsoft Excel and statistical Package for the Social Sciences (SPSS). In cases where the researchers did not receive a recording of an intervention session, regardless of the reason, it was considered as if the session did not take place (i.e., no dosage for that day). This approach was applied irrespective of whether the parent reported that the intervention had occurred. No adjustments were made for missing data in the analyses. While ratings for stuttering outcomes and adherence were made shortly after the recordings were received by researchers, ratings for quality and responsiveness were made four to six years later due to the need for additional training and calibration of raters to ensure consistency and accuracy. Despite this delay, all recordings were available and utilized for the evaluation. During this period, some recordings had become unusable, for example a number of the video files were corrupted. As a result, 4.2% of the recordings were missing from the analysis for quality and responsiveness.

The first research question was, what is the relationship between each FOI measurement (dosage, adherence, quality, and responsiveness) and each outcome measurement (%SS, SPM, naturalness, and severity) for children participating in the STOC pilot study? To address the first research question, Pearson Product-Moment correlations were conducted between each FOI variable (dosage, adherence, quality, responsiveness) and each outcome measurement (%SS, SPM, naturalness, severity). In instances where the correlation was derived from repeated measures within subjects, leading to autocorrelation, the coefficients were adjusted using multiple regression methods (see Bland & Altman, 1995). The strength of correlation was interpreted as weak (.0-.3), moderate (.3-.7), or strong (.7-1.0) (Ratner, 2009), with significance set at $p < .05$ for two-tailed comparisons. Given the high number of correlation coefficients, the Benjamini-Hochberg (1995) procedure for controlling the false discovery rate (FDR) was utilized. Twenty-eight correlations were performed (seven children, twice daily sessions for four phases [A_1, B_1, A_2, B_2]). The second research question was, do differences exist in the FOI measurements between participants who completed the STOC intervention program and those who did not? (c) What level of fidelity was achieved in the STOC study? To address the second research question, analysis of mean differences confidence intervals (CI) was performed.

3.3 A pilot study of stuttering treatment for older children (Study II)

In this study, seven participants' changes in fluency were documented based on an examination of performance in both intervention sessions and monthly recordings. The intervention yielded promising outcomes across participants, demonstrating reductions in both the frequency (%SS) and severity of stuttering while also enhancing syllables per minute (SPM) and naturalness. Notably, the four participants who successfully finished the STOC program reported improvement in quality of life. For the three participants who withdrew from the program, the data is only available after intervention (B_2). The main reasons given for withdrawing was lack of motivation to persist with the program.

To evaluate the effectiveness of the STOC intervention, the %SS measures from monthly sessions were assessed. In all the figures that follow, the phases labelled A are periods in which no intervention occurred and phases labelled B were periods of intervention (see Figure 4). Figures 5-11 show each participant's monthly measurement data for each of the four outcome measures (%SS, SPM, naturalness rating, and severity rating). Graphs displaying outcome data for each intervention session for each participant are presented in Appendix D.

Figure 5. Graph of Outcomes Variables for P1

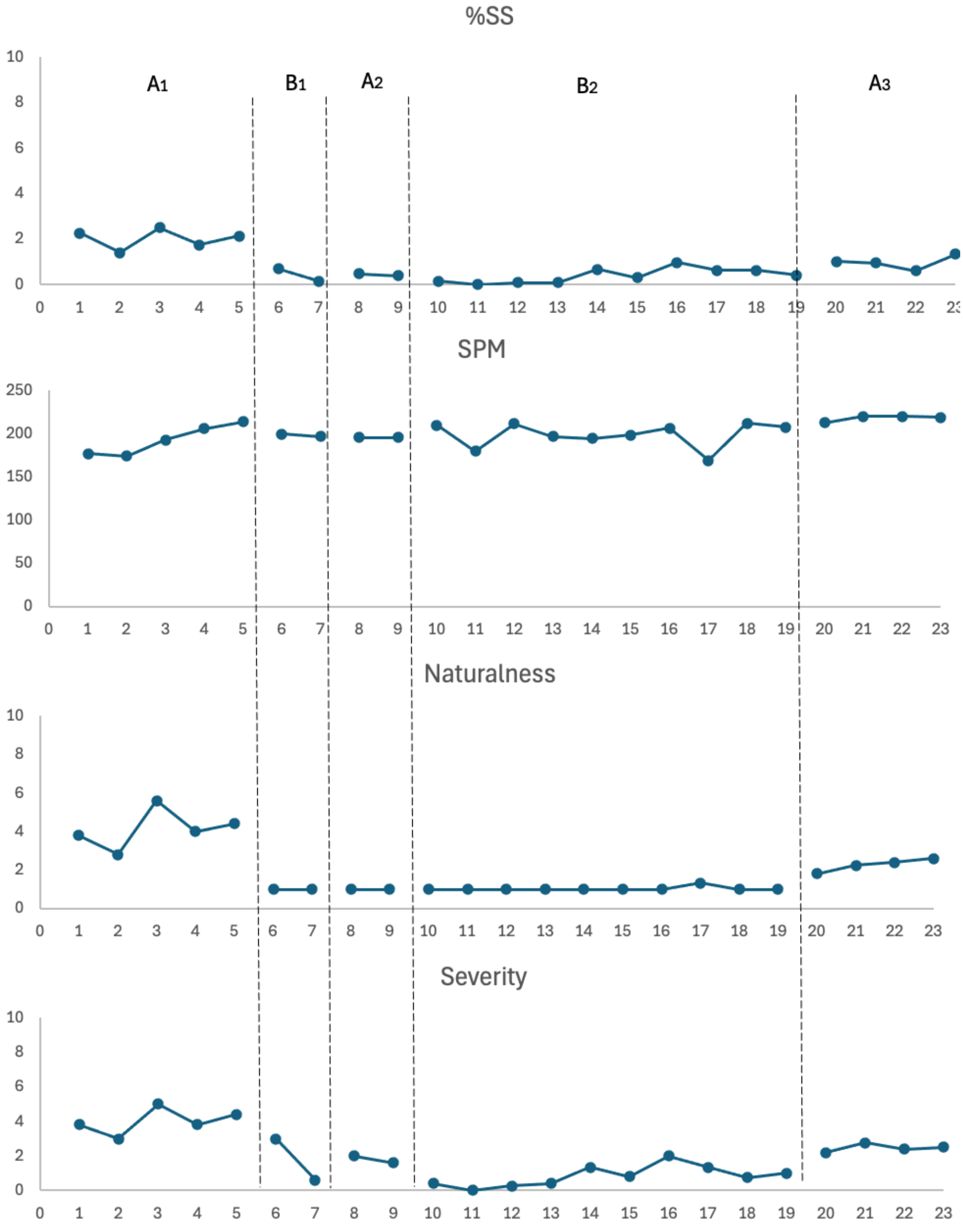


Figure 6. Graph of Outcomes Variables for P2

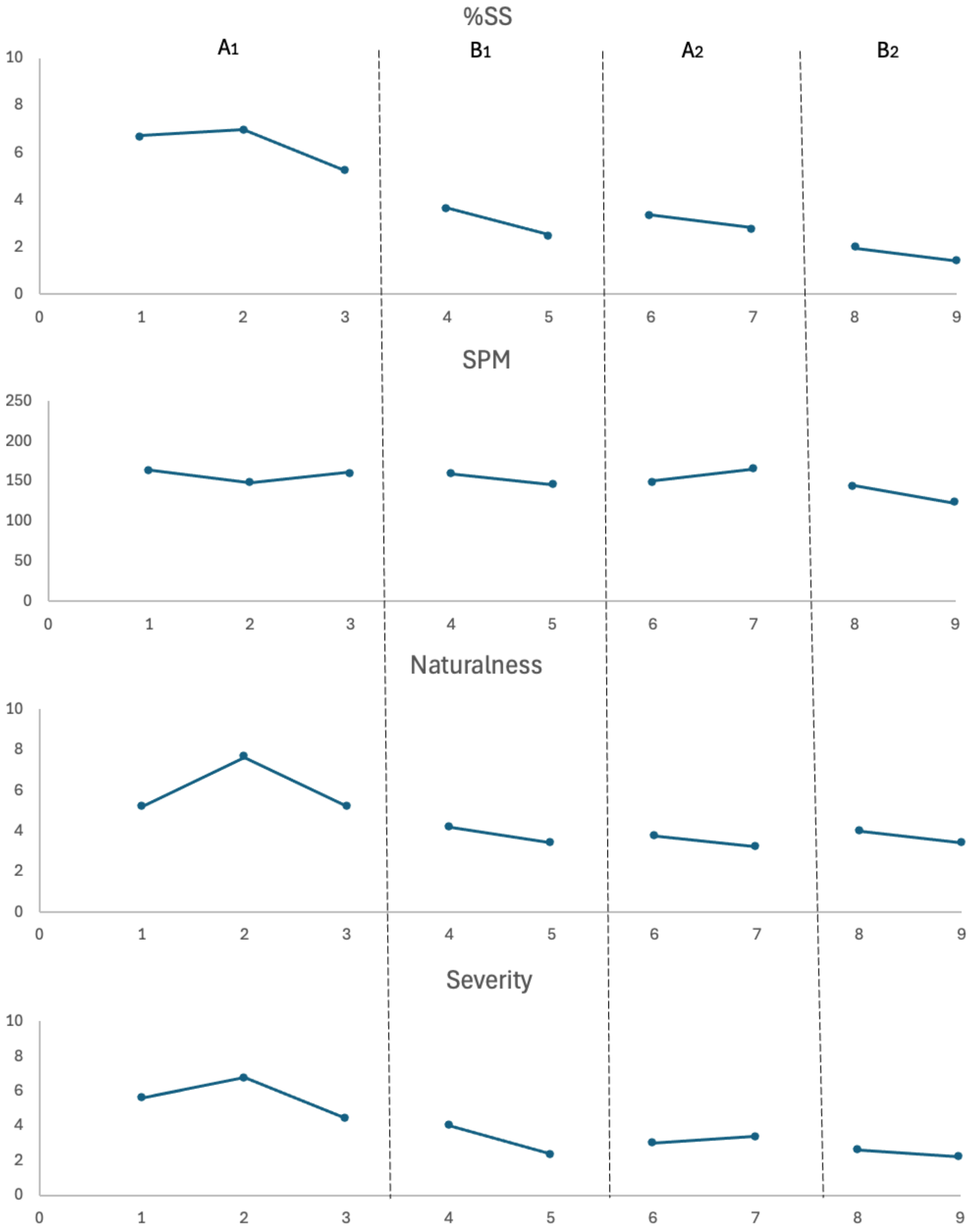


Figure 7. Graph of Outcomes Variables for P3

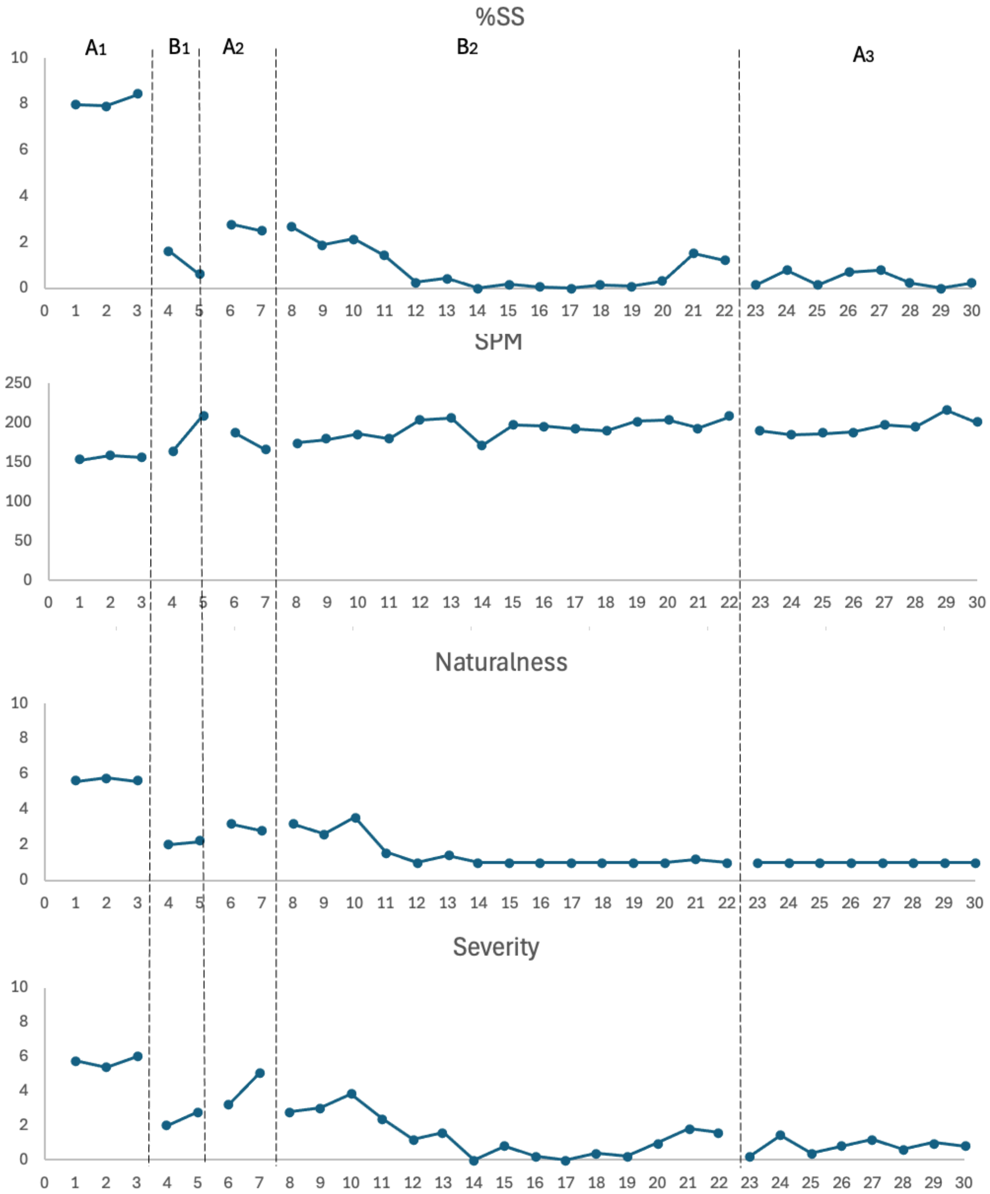


Figure 8. Graph of Outcomes Variables for P4

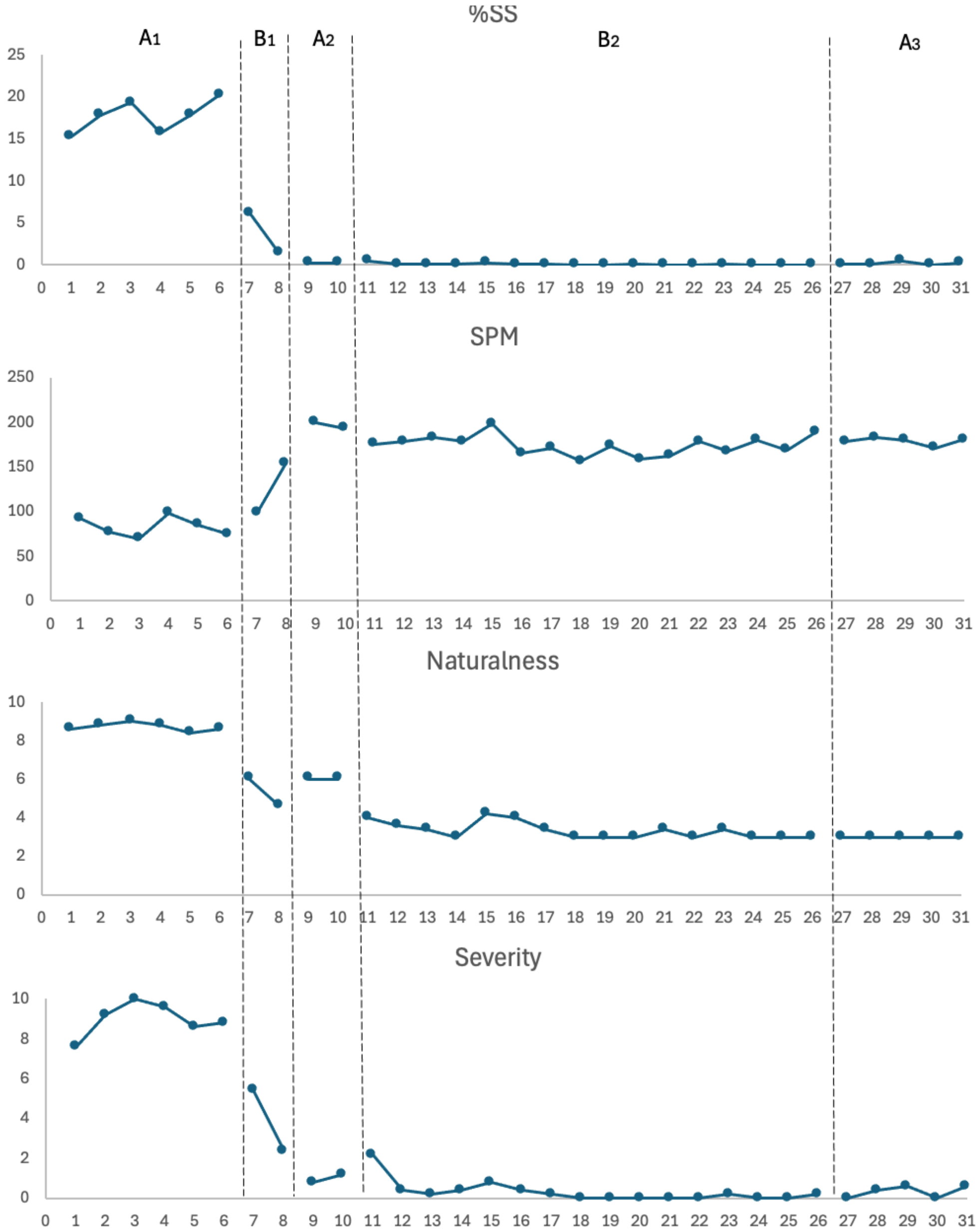


Figure 9. Graph of Outcomes Variables for P5

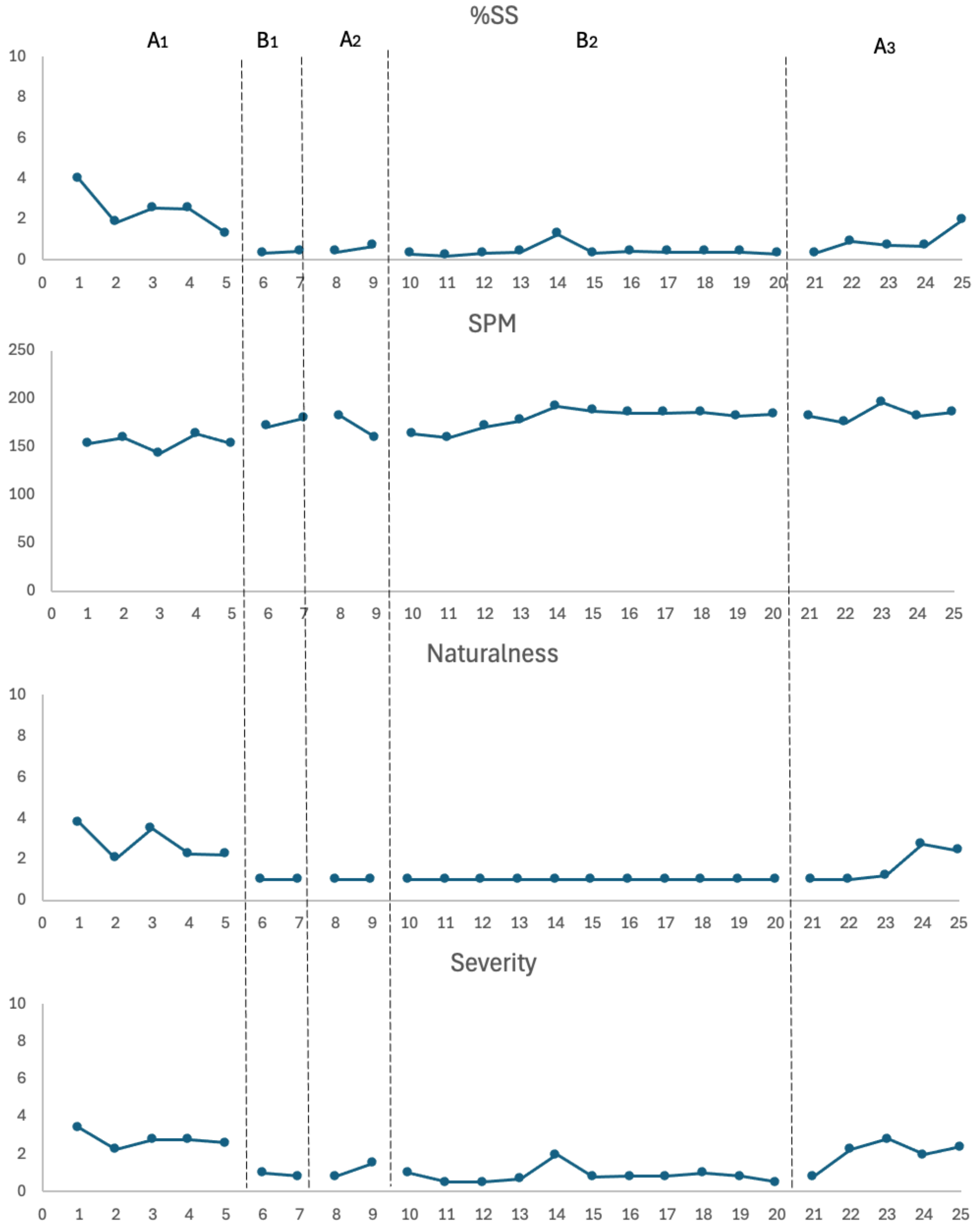


Figure 10. Graph of Outcomes Variables for P6

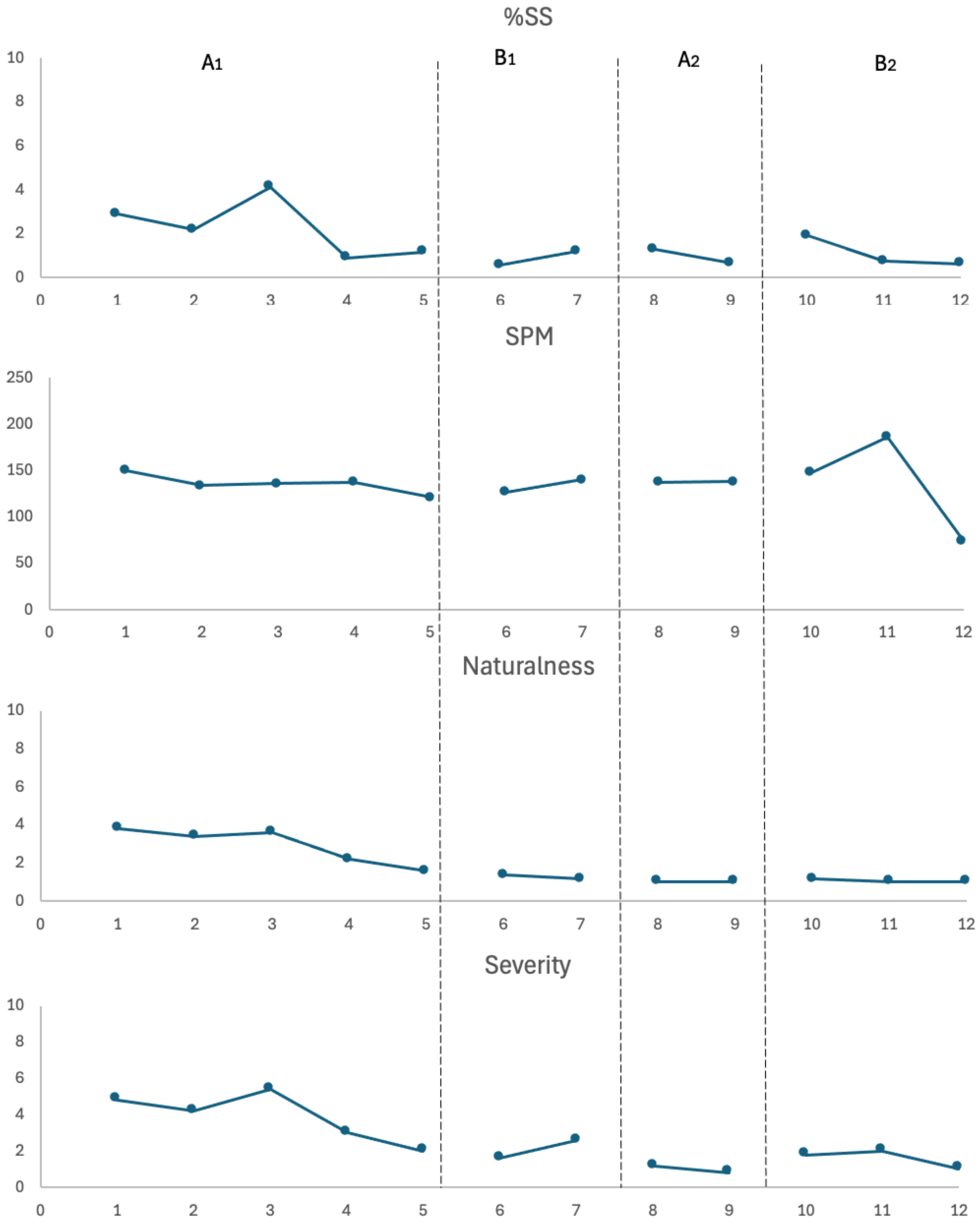


Figure 11. Graph of Outcomes Variables for P7

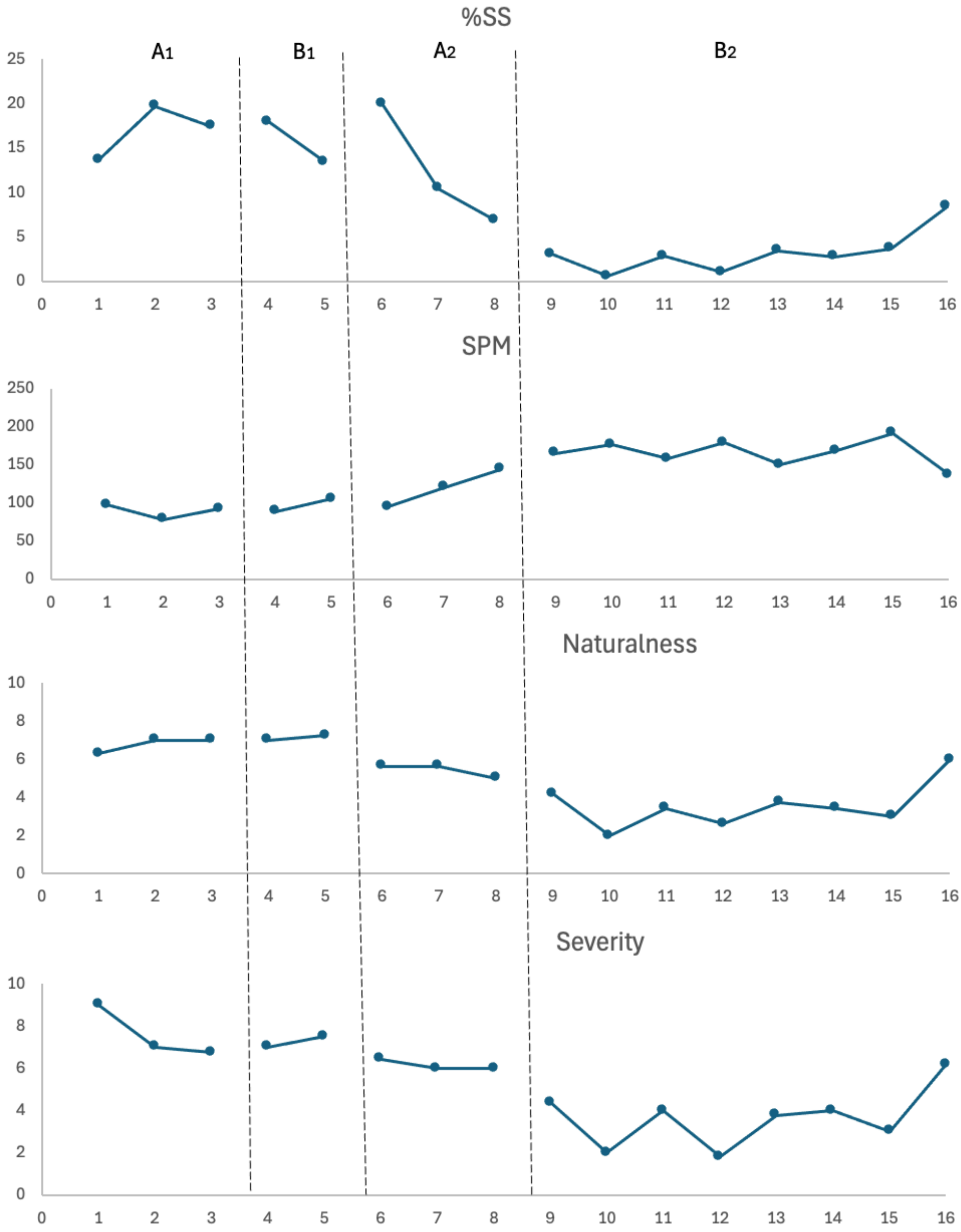


Figure 12 illustrates %SS measures in monthly sessions for all participants. These values represent averages for each phase of the study, including follow-up. As can be seen in Figure 12, across the study all participants experienced a reduction in stuttering. The variation in %SS between participants at the first monthly measurements was substantial (*Range*=2.0-17.7). As the %SS reflects the frequency of stuttering during speech, the range indicates a notable disparity among participants regarding the severity of stuttering at the beginning of the intervention. It is evident that the %SS decreases for all participants, albeit at varying phases for each participant. Particularly noteworthy was the substantial reduction in stuttering observed for participants P3 (96.3%) and P4 (98.9%), both of whom exhibited more severe stuttering during the baseline phase than the other participants. Participant P7 also had severe stuttering during baseline but the reduction of %SS was slower to happen in his case (77.6%). Stuttering persisted at a reduced rate for all participants throughout the study, including in B1, A2, B2, and during follow-up (A3). However, for P3, stuttering increased during the second baseline (A2) before declining again after four months of intervention (B2). Conversely, the results for participants who did not complete the program are less definitive. Nonetheless, stuttering decreased from the first baseline (A1) in subsequent conditions for the three participants who did not complete the program, indicating an effect of STOC. A Related-Samples Friedman's non-parametric ranks test was applied to the percentage of syllables stuttered from the monthly recordings. The mean percentages for each condition across all seven participants were used. Follow-up data (A_3) was unavailable for the three participants who did not complete the program. The analysis revealed a significant difference in the percentage of syllables stuttered between conditions, with Friedman's test statistic (4) =10.6, *p*=.016 (α=.05; one-tailed test). Pairwise comparisons indicated that the significant differences occurred between the initial baseline (A_1) and each of the subsequent conditions.

Figure 12. %SS Measurements in Monthly Sessions

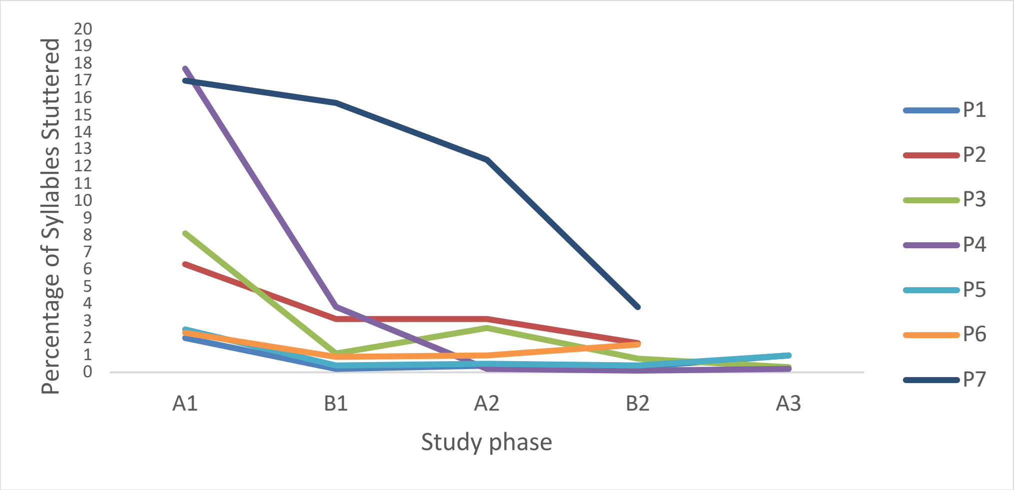
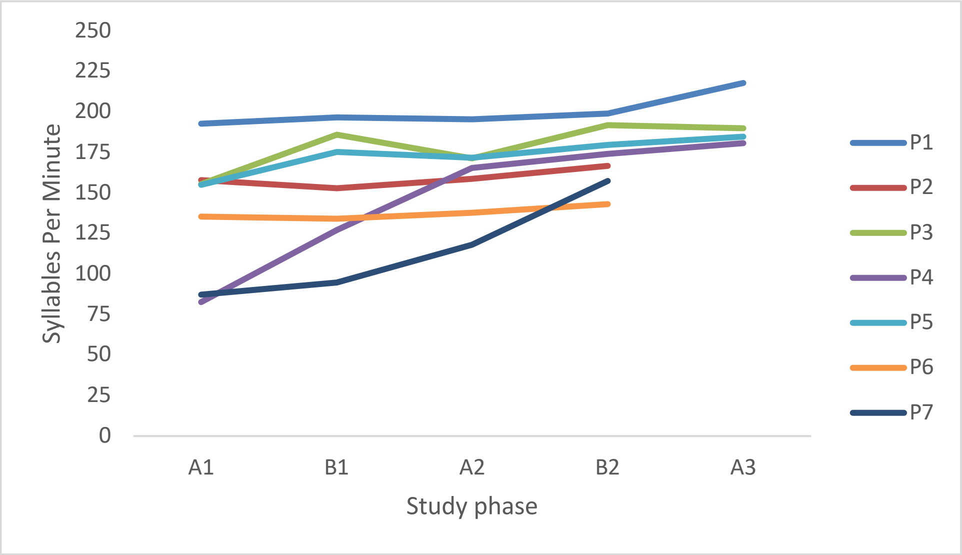


Figure 13 illustrates Syllables per minute (SPM) measures in monthly sessions for all participants. The variation in SPM between participants at the first monthly measurements was substantial (*Range*=82.6-192.6). As the SPM reflects the number of syllables spoken, the range indicates a difference among participants regarding the amount of speech produced during the three minutes of speaking time assessed. It is evident that the SPM increased for all participants, albeit at varying phases for each participant. Particularly noteworthy was the substantial increase in SPM observed for participants P4 (118.6%) and P7 (80.4%), both of whom exhibited lower SPM during the baseline phase than the other participants. SPM increased for all participants throughout the study, with some variations. SPM was lower after 2 months of intervention (B_1) for participants P2 (decrease of 3.1%) and P6 (decrease of 1.1%). The results for participants who did not complete the program are less definitive.

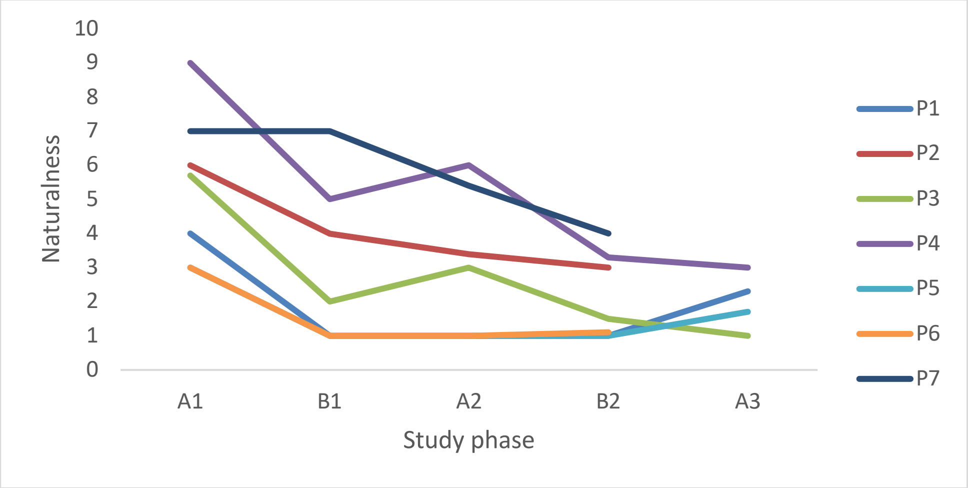
Figure 13. SPM Measurements in Monthly Sessions



Nevertheless, SPM increased from the first baseline (A_1) in subsequent conditions for the three participants who did not complete the program, indicating an effect of STOC. A significant difference between conditions was identified (*Friedman's test statistic* (4)=11.8, *p*=.01). Pairwise comparisons indicated that syllables per minute were notably higher in B_2 and A_3 (follow-up) compared to the initial baseline condition (A_1).

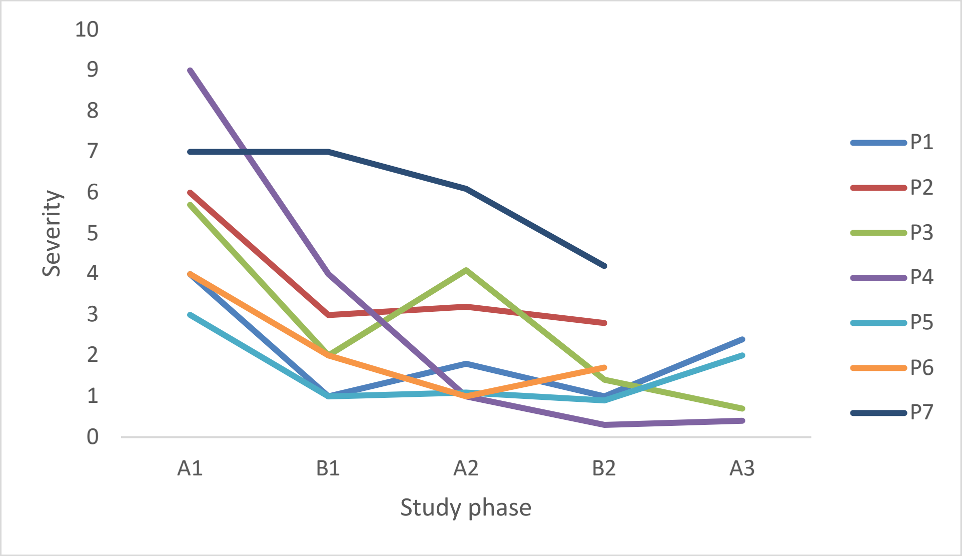
Figure 14 illustrates naturalness measures in monthly sessions for all participants. The variation in naturalness between participants at the first monthly measurements was considerable (*Range*=3-9). As the naturalness reflects how natural the participants spoke, the range indicates a difference among participants. Naturalness was rated from 1 (highly natural) to 9 (highly unnatural). It is evident that the participants speech became more natural, albeit at varying phases for each participant and with variations at the end. It is evident that from first baseline (A_1) all participants improved their naturalness, with the exception of P7 where the naturalness does not improve until the second baseline (A2). The improvement in naturalness is noteworthy for all participants, but as only four participants went on to the follow-up (A3) where there was an increase for two of the participants, it is hard to evaluate the outcome for the three participants who did not complete the program. The results indicate however a positive effect of STOC towards more natural sounding speech. A notable difference in naturalness was observed between conditions (*Friedman's test statistic* (4)=7.9, *p*=.024). Pairwise comparisons indicated that naturalness was significantly higher (lower number) in B_1 and B_2 compared to A1.

Figure 14. Naturalness Measures in Monthly Sessions



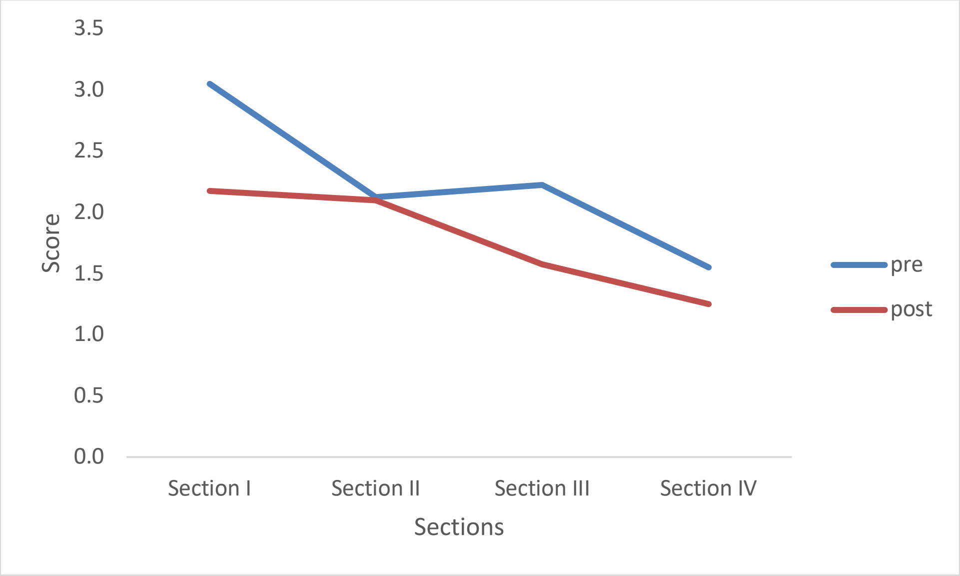
Changes in the outcome measure of severity were also observed across the phases of the study (see Figure 15). The variation in severity between participants at the first monthly measurements was (*Range*=3-9). The severity scale ranges from 0 to 10, zero indicating no stuttering and therefore no severity, 10 is rated to be extremely severe stuttering. The severity scale indicates the overall severity of the stutter indicating a notable disparity among participants at the beginning of the intervention. It is evident that the severity decreased for all participants, albeit at different phases for each participant. Particularly noteworthy was the substantial reduction in severity for participant P4 who was given a severity rating 9 at the first baseline (A1) which reduced to 0.4 at the follow-up phase (A3). All participants exhibited a reduction in the rating of severity of stuttering between first baseline (A1) and two months of intervention (B2), with the exception of P7 who had no decrease at that stage. The improvements in naturalness are noteworthy for all participants. However, since only four participants proceeded to the follow-up (A_3), where two showed further improvement, it is difficult to evaluate the outcome for the three participants who did not complete the program. Nonetheless, the results suggest a positive effect of STOC towards less sever stuttering. A notable difference in severity was observed between conditions (*Friedman's test statistic* (4)=9.3, *p*=.013). Pairwise comparisons revealed that severity was significantly lower in A_2 and B_2 compared to A_1.

Figure 15. Severity Measures in Monthly Sessions



The four participants who completed the program and follow-up filled out the OASES-S and the SPQ questionnaire. All participants reported improvements in speech fluency on the SPQ after intervention and indicated they could control their speech more effectively. They mentioned they were always or almost always able to speak without thinking about controlling their speech, a change from their pre-treatment experience. All participants attributed their speech improvements to the STOC intervention; however, only one participant reported being completely free of stuttering. Figure 16 shows the average measurement between sections in the OASES-S questionnaire for both pre intervention and post intervention. Section I covers general information about speech and fluency, Section II addresses reactions and feelings towards stuttering, Section III pertains to communication in daily situations, and Section IV involves questions about quality of life. A lower score in Figure 16 indicates a higher quality of life. The only sub-test of the OASES-S on which participants did not show improvement was Section II. On average the results showed a positive impact from the STOC intervention.

Figure 16. Average Measurement of OASES-S Scores



3.4 Results of evaluation of fidelity measurements (Study III)

The FOI measurement results are presented in detail in Paper III. A total of 1551 intervention session were delivered throughout the intervention. As can be seen in Table 4, the average for the FOI components across participants was higher for the parent-managed sessions than for the child-managed sessions for both dosage and adherence. Dosage was the FOI component that was most often achieved, but as can be seen in Table 4 the range for each component vary, especially for dosage and adherence.

Table 4. Results of the Average FOI for Parent- and Child-managed Intervention Sessions

	Parent	Child	P/C together	Range
Dosage	85.9%	80.5%	83.2%	58.5–97.7%
Adherence	64.6%	53.1%	58.9%	12.8–80.9%
Quality	1.7	1.6	1.7	1.0–2.2
Responsiveness	1.1	1.1	1.1	1.0–1.4

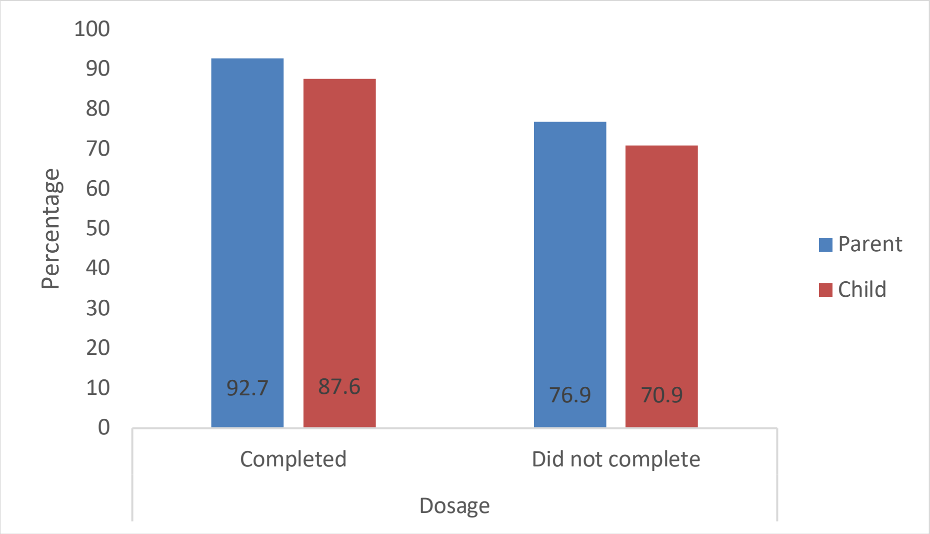
To assess the impact of each FOI component on the outcome measures a Person Product-Moment correlation coefficients were computed. The results indicated that no significant coefficient was found between dosage and any outcome measure. However, 5 out of 12 comparisons show moderate strength ($r < .29$), but low N resulted in none being significant. Negative correlations with %SS indicated that a higher dosage was associated with reduced stuttering, while positive correlations with SPM showed that a higher dosage was linked to increased speech output. Additionally, negative correlations for naturalness and severity suggested that speech became more natural and stuttering less severe with higher dosage. The correlation coefficients between dosage and severity were moderately strong, while those with naturalness were close to zero. In sessions managed by children, %SS, SPM, and severity showed moderate correlations with dosage (See Table 3 in Paper IV).

In the parent-managed sessions, 17 out of 28 correlation coefficients were significant between adherence and the four outcome measures, compared to 16 out of 28 in child-managed sessions. While most coefficients were weak, 11 out of 28 were of moderate strength in the parent-managed sessions, and 12 out of 28 were of moderate strength in the child-managed sessions, with one being strong. Negative coefficients with %SS indicated that higher adherence was associated with reduced stuttering, and negative coefficients with SPM indicated that higher adherence was associated with decreased speech, though only two coefficients were significant. For naturalness and severity, negative coefficients suggested that higher adherence led to more natural speech and less severe stuttering. In both parent- and child-managed sessions, most correlation coefficients between adherence and %SS were moderately strong, whereas those between adherence and SPM were generally weak. In the child-managed sessions, three coefficients between adherence and naturalness showed moderate strength, and one indicated a strong correlation (see Table 4 in Paper IV).

In the parent-managed sessions, 20 out of 28 correlation coefficients were significant between quality and the outcome measures, while 12 out of 28 were significant in child-managed sessions. Most correlations were weak, with 8 out of 28 showing moderate strength and one strong correlation in parent-managed sessions. Similarly, 8 out of 28 coefficients in child-managed sessions were of moderate strength. Negative correlations with %SS indicated that higher quality was associated with less stuttering, though this was significant in only 3 out of 14 cases. Negative correlations with SPM suggested that higher quality was linked to increased speech output, with most coefficients being significant. For naturalness and severity, negative correlations indicated that a higher quality of intervention was associated with more natural-sounding speech and less severe stuttering, with 10 out of 28 coefficients being significant. The majority of correlation coefficients between quality and SPM were moderately strong (8 out of 14), with one being strong. In contrast, most correlations between quality and %SS were weak (9 out of 14). Similarly, most correlation coefficients between quality and naturalness (8 out of 14) and between quality and severity (9 out of 14) were also weak (See Table 5 in Paper IV).

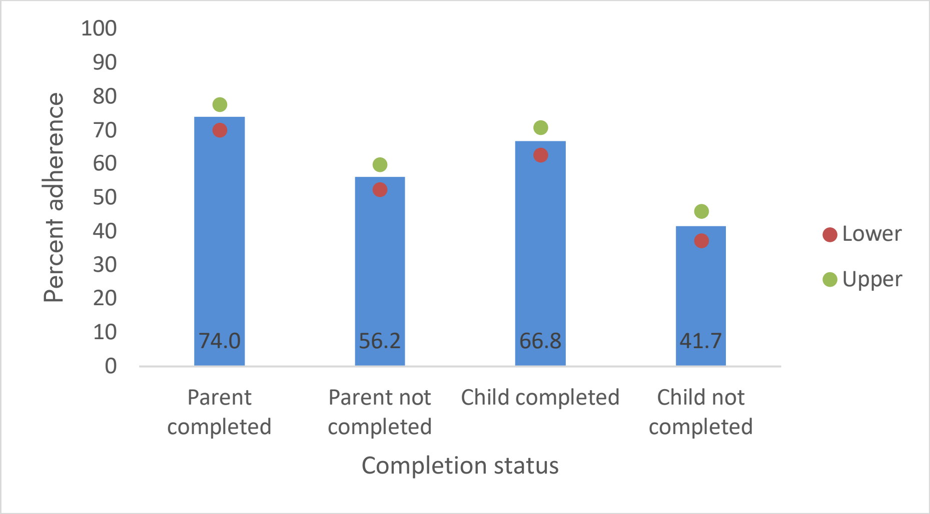
In the parent-managed sessions, 12 out of 28 correlation coefficients were significant between responsiveness and the outcome measures, compared to only 8 out of 28 in child-managed sessions. Most coefficients were weak, with just 3 of 28 showing moderate strength in parent-managed sessions and 1 of 28 in child-managed sessions. Three significant negative correlations between responsiveness and %SS were identified, indicating that higher responsiveness was associated with reduced stuttering in these cases. Additionally, significant negative coefficients were identified with SPM in half of the cases, suggesting that higher responsiveness was linked to increased speech output. For naturalness and severity, both positive and negative correlations were found, varying by participant. A positive correlation indicates that high responsiveness, exemplified by engaging and positive conversations, was associated with more natural-sounding speech and less severe stuttering. However, most coefficients were weak, with about half being 0.10 or lower, suggesting that responsiveness was generally not strongly related to naturalness and severity for most participants (See Table 6 in Paper IV).

Figure 17. Mean Difference in Dosage Based on Completion Status



To evaluate if there was a difference between the participants that finished the STOC intervention (P1, P3, P4, P5) and those who did not (P2, P6, P7), FOI results were compared for these groups. As can be seen in Figure 18, the mean was higher for dosage in the group that finished the intervention, both for the parent-managed and the child-managed sessions.

Figure 18. Mean and Confidence Intervals for Adherence Based on Completion Status



To evaluate the differences in adherence, responsiveness, and quality for the two groups (completed STOC, did not complete STOC), an analysis of mean differences was conducted by employing confidence intervals (CI). As can be seen in Figure 18, participants who completed the STOC intervention, had higher adherence in both parent-managed sessions (*M*=74.0%) and child-managed sessions (*M*=66.8%) compared to those who did not complete the STOC intervention. For adherence, the absence of overlap between the group that completed the STOC intervention in parent-managed sessions, 95% CI [70.1, 77.7] and the group that did not complete the STOC intervention, 95% CI [52.4, 60.0] suggests that the difference is unlikely to be due to chance. Likewise, for the group that completed the STOC intervention in the child-managed session 95% CI [62.7, 70.9] compared to the group that did not complete the STOC intervention 95% CI [37.3, 46.0] the absence of overlap suggests that the difference is unlikely to be due to change.

When looking at results for quality, 1 is the highest quality “the conversation is flowing and the child responds with full-length answers” and 4 is the lowest quality, “the child is not taking part in the conversation (mostly yes and no answers)”. As seen in Figure 19, participants who completed the intervention, the quality scores were better for both parent-managed sessions (*M*=1.46) and child-managed sessions (*M*=1.38) compared to those who did not complete the intervention.

For quality, the confidence intervals for the group that completed the STOC intervention in parent-managed sessions 95% CI [1.40, 1.78] and the group that did not complete the STOC intervention 95% CI [1.78, 1.96] are close but non-overlapping, suggesting that this difference is unlikely to be due to chance. The lack of overlap in the confidence intervals for the group that completed the STOC intervention in child-managed sessions 95% CI [1.32, 1.45] compared to the group that did not complete the STOC intervention 95% CI [1.66, 1.84] indicates that this difference is unlikely to be coincidental. The lack of overlap in the confidence intervals, upper and lower bond, indicates that this difference is unlikely to be coincidental (see Figure 19).

For responsiveness, 1 was the highest rating "very happy - conversation good" and 3 was the lowest "not happy to take part In the Intervention/conversation. Responsiveness was the only component out of the four FOI components were no difference between the groups who completed and did not complete the intervention. As seen in Figure 20, the means for participants who completed the intervention (parent-managed *M*=1.12; child-managed *M*=1.07) and did not complete the intervention (parent-managed *M*=1.13; child-managed *M*=1.12) as similar. As can be seen in Figure 20, for responsiveness the near-complete overlap of the confidence intervals of the group that completed the STOC intervention in parent-managed sessions 95% CI [1.09, 1.15] and the group that did not complete the STOC intervention 95% CI [1.09, 1.17] suggest there was no significant difference between the groups. This near-complete overlap of the confidence intervals can also be seen in the child-managed sessions for the group that completed the sessions 95% CI [1.04, 1.10] and the group that did not complete the sessions 95% CI [1.08, 1.16].

Figure 19. Mean and Confidence Intervals for Quality Based on Completion Status

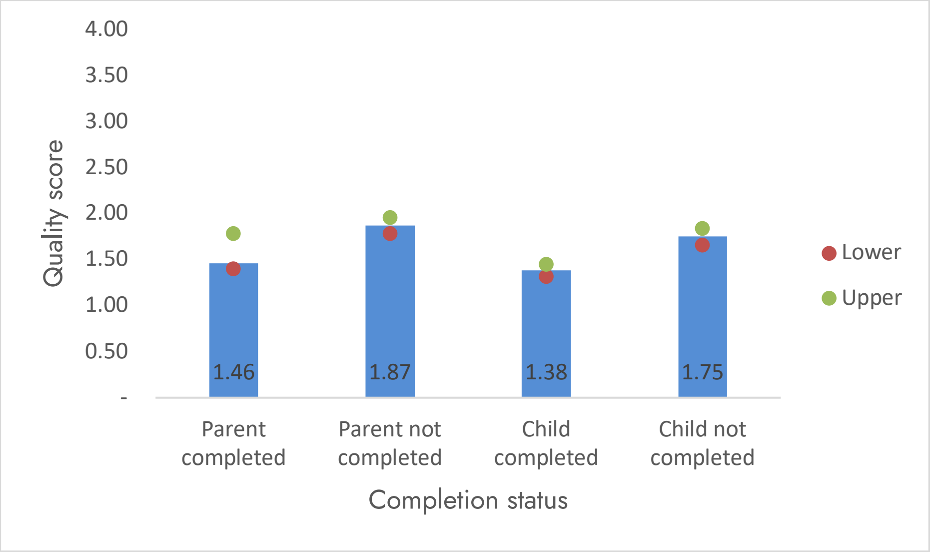
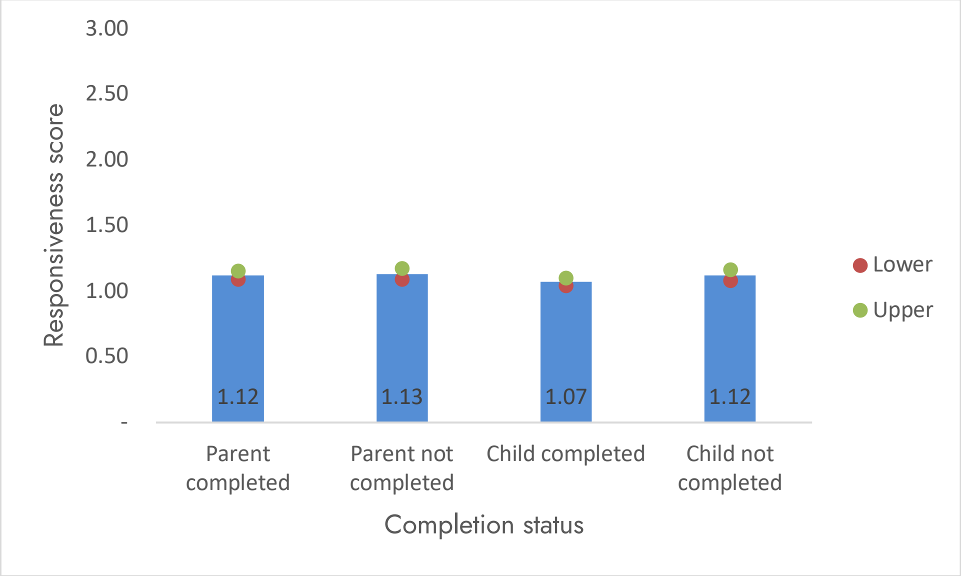


Figure 20. Mean and Confidence Intervals for Responsiveness Based on Completion Status



4 Discussion

4.1 Study I

The primary objective of this study was to evaluate the reporting of FOI components in research on parent-implemented interventions for CWS (Bergþórsdóttir et al., 2022). The focus was specifically on interventions that included a home-based implementation. By systematically reviewing the existing literature, we sought to identify patterns and standards in the reporting practices of FOI measurements. Additionally, we aimed to pinpoint gaps in the current knowledge base to highlight areas where further research is needed. This research provides critical insights into the effectiveness and reporting quality of home-based stuttering interventions, ultimately guiding future research, and improving intervention strategies for CWS. This study identified that FOI components were more frequently addressed in clinical settings compared to home settings. However, overall, the reporting of FOI was limited in both contexts. The reporting of FOI components rarely included FOI data, except for dosage in clinical settings. In clinical settings, dosage was the most commonly reported FOI component, followed by responsiveness, adherence, and quality. Conversely, in home settings, adherence and quality were the most frequently reported FOI components, with dosage and responsiveness following behind. Notably, the robustness of FOI reporting was demonstrated through the inclusion of evidence of dosage measurement, both in clinical and home settings. Other FOI components were typically reported based on narrative accounts and occasionally presented ambiguously. Nevertheless, most FOI components were scarcely reported. It was anticipated that there would be an increase in FOI reporting over time; however, there was only a weak trend for improvement, with no significant change observed.

These findings highlight the need for further research and standardization in the integration of FOI components into stuttering interventions to optimize therapeutic outcomes in both clinical and home environments and to develop a more robust evidence-base for clinical practice. The contribution to the existing literature on FOI in intervention studies, focusing on this new population group of CWS, align with previous research highlighting the generally poor FOI of parent-implemented intervention for children with differences and disabilities (Dane & Schneider, 1998; Gresham et al., 1993; Hofslundsengen et al.,2022; Moncher & Prinz, 1991; Perepletchikova et al., 2007). Notably, differences were observed between this study and Lieberman-Betz's (2015) research on FOI in studies of interventions for children's language skills. In both studies, the most frequently reported FOI was dosage for clinicians and adherence for parents. Comparable rates were observed for various FOI components. In clinic-based dosage, the rates were nearly identical between Lieberman-Betz's study, which reported 71%, and the present study, which found 72.2%.

Similarly, while both studies reported on home-based dosage rates, there was a noticeable difference, with Lieberman-Betz reporting 14% and the present study observing 19.4%.

The quality of interventions in clinical settings showed similar findings, with Lieberman-Betz at 20% and the present study at 19.4%, and at home, quality rates were also comparable, with 29% in Lieberman-Betz's study and 30.6% in the present study. However, the present study reported notably lower rates for certain FOI components. Specifically, adherence in clinical settings was significantly lower in the present study at 19.4% compared to Lieberman-Betz's 34%. Similarly, adherence at home was lower, with the present study reporting 36.1% versus Lieberman-Betz's 60%. For responsiveness in clinical settings, the present study found a rate of 30.6%, whereas Lieberman-Betz reported 54%. Likewise, responsiveness at home was lower in the present study at 19.4% compared to Lieberman-Betz's 43%. The reason for these differences remains unclear. The number of studies was similar (35 in the Lieberman-Betz study and 36 in the present study). The age groups in the studies were also comparable, with Lieberman-Betz focusing on children aged 0-6 years and the present study on children aged 2-6 years. Additionally, the years of the studies under examination were from 1988-2012 for Lieberman-Betz and from 1990-2020 for the present study. Lieberman-Betz (2015) identified a trend of increasing FOI reporting over time in their data, a trend that was also observed in the present study, although to a lesser degree.

Despite an increase in the number of studies published on parent-implemented interventions for CWS, there has not been a corresponding rise in the reporting of FOI components. This gap is problematic as inadequate FOI reporting hampers the systematic evaluation of intervention efficacy, leaving clinicians and researchers without key information to explain variations in outcomes across studies. Although authors of all 36 included papers included in Study I provided relatively clear descriptions of research designs, protocols, and outcome measures, the lack of detailed FOI reporting prevents a thorough understanding of what was actually implemented according to protocol. This issue is consistent with findings from systematic reviews by Brignell et al. (2021), Hofslundsengen et al. (2022) and Johnson et al. (2023) who reported that while intervention designs are generally well-reported, reporting of FOI components were often overlooked. Brignell et al. (2021) highlighted the need for research to address the relationship between dose and response, and Nye and Hahs-Vaughn (2011) emphasized the importance of transparency and detailed reporting in intervention research to facilitate replication and validate findings. This also affects the potential for exportability and dissemination (Borrelli et al., 2005).

Given the growing emphasis on FOI measurement in recent research (An et al., 2020; Hofslundsengen et al., 2022), it is surprising that many newer intervention studies have not fully incorporated these measures into their designs. Reviewing recent studies on stuttering intervention, that were published after the systematic search for Study I was completed, poor reporting of FOI is still evident. Druker et al. (2020) conducted an intervention on evidence-based parenting support for treating early developmental stuttering, involving 77 children aged between 2.0 and 5.3 years. In their study, FOI was limited to reporting dosage and responsiveness in the clinical setting, with responsiveness being the only component measured in the home setting. Similarly, Euler et al. (2021) carried out a group intervention focused on speech restructuring for children aged six to nine, where the focus was on reinforcing soft voice onset. While their study noted that children and parents participated in various intervention activities together, there was no mention of how these activities were measured or evaluated, and no FOI components were reported in their results. Preston et al. (2022) described Palin parent-child interaction therapy for children with autism spectrum disorder who stutter, with three children aged 4.5, 6.7, and 7.7 years. Their study only reported the FOI component of dosage in the clinical setting and responsiveness in the home setting. Meanwhile, Johnson et al. (2024) examined the Lidcombe Program delivered as a telehealth intervention, involving 37 CWS aged between six and twelve years. In their design, parents were required to submit recordings of intervention sessions, and the responsiveness for both clinical and home settings was reported in their paper.

In conclusion, despite the growing emphasis on FOI in intervention research, changes in FOI measurement and reporting within the literature concerning interventions for CWS remains minimal (Bergþórsdóttir et al., 2022). While the design of interventions is often well-described, it remains unclear whether the interventions have been executed according to protocol. One might expect that recent studies would integrate FOI measures more rigorously into their designs, yet this has not been consistently observed. Had we considered other aspects of FOI, such as those outlined in the NIH BCC treatment fidelity framework Bellg et al., 2004; Borrelli et al., 2005), the results might have been different. For example, incorporating the training domain, which focuses on the competencies required for successful delivery and the standardization of training, might have highlighted variations in provider skill acquisition and maintenance. Similarly, the delivery domain, which confirms that the intervention is delivered as intended and assesses nonspecific effects, could have revealed differences in therapeutic alliance. If these aspects had been included, the results might have shown greater variability, potentially highlighting areas where the intervention was less effective. This could change the interpretation of the intervention's overall effectiveness, as it would account for more nuanced factors influencing outcomes. To advance the field, clear guidelines are needed for stuttering intervention trials to facilitate the systematic and thorough collection and reporting of FOI. Such improvements are crucial for better understanding the processes through which interventions effect change and for enhancing the reliability and replicability of research outcomes.

4.2 Study II

The aim of this study was to examine the effectiveness of a pilot home-based intervention for CWS (Einarsdóttir et al., 2024). In this study of STOC, there was a specific focus on incorporating detailed measurements of FOI into the research design. By integrating FOI measurements, this study sought to provide a more accurate assessment of the intervention's impact on stuttering outcome measures. This approach not only allowed for a thorough evaluation of STOC's effectiveness, but also contributed to a deeper understanding of how home-based interventions can be optimized for CWS and how rigorous monitoring of FOI is possible in home-based interventions. The findings from this study are intended to bridge gaps in existing research, offer practical insights for clinicians, and guide future research endeavours in developing robust and effective stuttering interventions. STOC, building on the Response Contingent Time-Out principle, was designed as a parent-implemented program utilizing parent-imposed time-out and self-imposed time-out techniques as time-out interventions have consistently proven effective in treating childhood stuttering (for review see Brignell et al., 2021).This study demonstrated these methods to be effective in modifying speech and promoting increased fluency, especially in controlled situations. All participants experienced a reduction in stuttering frequency, an increase in the quantity of speech, more natural-sounding speech along with less severe stuttering. Individual responses to the program varied, with only four of the seven participants completing the entire intervention. Upon completion of the intervention participants reported how often they were able to speak fluently without thinking about modifying or controlling their speech. One reported that this was always the case, two reported this was almost always the case, and one reported that this was sometimes the case.

The STOC intervention features awareness training both for the parent and the child, where the time-out was used contingent on stuttering events (James et al., 1989). Studies on interventions using time-out have shown that participants respond differently to the intervention, with some responding better than others (e.g., Baxter et al., 2016; Mallick et al. 2021). In the STOC study, all seven participants increase in fluency during the intervention. However, three did not meet the fluency requirements to move to the next step of the intervention where the frequency of sessions was decreased contingent on fluent speech. This requirement was to have six consecutive sessions with less than 1% SS, a 1–3 rating on the naturalness scale, and a 0–2 rating on the severity scale. While the rate of children who failed to progress to the second intervention stage (B_2) in the STOC study seems high, the progress of participants is better than that in equivalent studies of other interventions. For example, in Johnson et al.'s (2024) trial of the Lidcombe Program with school-aged CWS 30 of the 37 (81.1%) children did not meet the fluency requirements to progress from Stage 1 (intervention) to Stage 2 (maintenance). In contrast, in the STOC study 100% of the children completed the first phase of intervention and first maintenance phases (B_1 and A_2). Various factors can account for the variation in individual responses, including challenges in effectively implementing the intervention techniques, which is important to monitor with FOI in terms of evaluating if the results are due to the implementation or the treatment itself (Smith et al., 2020). All participants in the present study demonstrated improved fluency following the intervention, with an average stuttering reduction of 72% from pre-treatment to follow-up. In contrast, more recent studies on school-aged CWS have shown lower overall reductions in stuttering after intervention: 57% in Euler et al. (2021), 67% in Johnson et al. (2024), and 25% in Kohmäscher et al. (2023).

The impact of STOC was evident not only in changing children’s fluency, but also in positively influencing children's attitudes toward their communication and stuttering in daily life. The decrease in OASES-S scores across all four sections mirrors findings from other recent studies on stuttering treatment for school-aged children (Euler et al., 2021; Johnson et al., 2024; Kohmäscher et al., 2023). From pre-treatment to the one-year follow-up, OASES-S scores showed a significant decrease, with a large effect size. This shift indicates that, relative to the baseline assessment, participants perceived notable progress, both in fluency and in quality of life (Euler et al., 2021). Johnson et al. (2024) results indicate a significant reduction between pre-treatment and 12 months post-treatment, changing from moderate to mild-moderate. At baseline, participants experienced a moderate level of stuttering impact, which subsequently shifted to a mild-to-moderate impact following the intervention (Kohmäscher et al., 2023). One contributing factor to this improvement in Study II could have been the open discussions between children and parents that the intervention encouraged. Throughout the intervention sessions, children were seen reminding their parents to use parent-implemented time-out and giving feedback on their accuracy in identifying stuttering moments. At the end of each session, both parties had to agree on whether a stuttering event occurred, knowing the researchers would provide feedback. These discussions may have reduced the children's anxiety around stuttering, aligning with Croft and Byrd (2020) findings on the benefits of open communication in reducing anxiety. Moreover, the focused attention that parents gave their children during these sessions could have contributed to the children's well-being. Thomas et al. (2020) found that parental involvement significantly influenced both the well-being and academic achievements of 7th grade students, suggesting that such engagement is a key factor in positive outcomes. In Study II, several parents informally reported to the SLP that STOC was associated with positive changes in their child's life outside the intervention setting. Previous research using self-imposed time-out with 22 participants older than 14 years who stuttered, the %SS was reduced by 50% with more than half of the participants. The results indicated that they were generally satisfied with both the intervention and their speech post-intervention, reporting that the treatment was easier to employ than other treatments they had engaged in (Hewat et al., 2006). By creating a space where stuttering events could be openly discussed and addressed, children may have improved their self-monitoring abilities and gained more confidence in their speech.

In the present study, all treatment sessions were recorded and evaluated by trained judges to assess FOI. After each session, both parents and children received feedback from the researchers on the frequency of stuttering and the accuracy of time-out implementation. This feedback was especially important during the initial weeks of treatment to confirm the intervention was carried out according to protocol and to maintain high fidelity. However, several challenges emerged in implementing the treatment effectively. For instance, one participant with ADHD struggled to consistently apply the time-out technique and pause during conversations, highlighting how neurodiversity can complicate intervention adherence. Additionally, the lack of family engagement and support in another participant’s home environment posed further challenges to achieving optimal fidelity in treatment delivery. These issues underscore the importance of addressing individual and contextual factors when implementing interventions to achieve effectiveness, such as ensuring that children and parents have the support and environment necessary to be able to implement an intervention with fidelity. Furthermore, monitoring FOI in both clinical and home settings is crucial, especially in interventions like this where parents play an active role (Hofslundsengen et al., 2022). FOI measurements in home settings help determine whether parents are accurately applying the intervention techniques outside the controlled clinical environment. This is important because home-based implementation can significantly influence the overall success of the intervention. Without proper fidelity in the home, the benefits seen in clinical settings may not fully transfer into the child's daily life, which could limit the intervention's long-term impact. Monitoring FOI in these settings can confirm that the intervention is being consistently applied, allowing for a more accurate assessment of its effectiveness and sustainability.

4.3 Study III

In this study, it was assessed whether high levels of fidelity could be achieved in an intervention where FOI was incorporated into the study design and closely monitored. In the context of a pilot study of the STOC intervention for school aged CWS, the relationship between four FOI components (dosage, adherence, quality, and responsiveness) and four outcome measurements (%SS, SPM, naturalness, and severity) was examined (Bergþórsdóttir et al., 2024). The findings suggest that dosage and adherence play big roles in reducing stuttering frequency, increasing speech fluency and mitigating severity. Increased adherence was associated with decreased stuttering frequency, indicating that effective utilization of the intervention led to tangible improvements. However, adherence did not correlate with an overall increase in the amount of speech produced by the children. Although better quality and responsiveness were not directly associated with changes in stuttering frequency or severity, higher quality and responsiveness allowed for more free flowing parent-child communication where the child had more opportunities to speak, regardless of the frequency or severity of stuttering within a session. In comparing parent-managed and child-managed intervention sessions, dosage and adherence were notably higher in parent-managed sessions. Interestingly, participants who completed the intervention exhibited higher FOI measurements compared to those who did not. This shows a positive relationship between adherence to intervention protocols and intervention effectiveness (not ruling out the possibility that other variables may be mediating this relationship).

These findings add to the existing FOI literature and highlight the critical need for incorporating FOI measurements in intervention studies. (An et al., 2020; Dane & Schneider, 1998; Hofslundsengen et al., 2022; Perepletchikova et al., 2007). The findings indicate that dosage plays a crucial role in reducing stuttering, increasing speech fluency and decreasing severity. This outcome aligns with the longstanding recognition of dosage as a key component of FOI across various types of behavioural interventions, such as drug abuse prevention (Dusenbury et al., 2003), early communication intervention (Lieberman-Betz, 2015), and psychotherapy (Perepletchikova et al., 2007). Most research on stuttering interventions includes some form of dosage reporting. Nye and Hahs-Vaughn's (2011) analysis of the methodological quality of experimental and quasi-experimental group-design studies in stuttering intervention highlighted significant gaps in the clarity and completeness of reporting on external and internal validity. These limitations complicate the interpretation of findings from individual studies, making it challenging to fully assess the reliability and applicability of their results. One of the key findings in Brignell et al. (2021) results in a systematic review on interventions for children and adolescents who stutter, were that studies should address the current gaps in understanding how to deliver interventions most effectively by exploring the relationship between intervention dose and response. A review of thirty-six stuttering intervention studies revealed that most provided limited or unclear information on whether and how FOI was monitored and measured (Bergþórsdóttir et al., 2022). The findings of a narrative systematic review, with the aim to assess the implementation quality of previously published group comparison clinical trials on stuttering interventions for children and adolescents (under 18 years), emphasize the need for future clinical trials on stuttering interventions to rigorously follow systematic guidelines for reporting implementation quality, ensuring the reliability of trial outcomes (Hofslundsengen et al., 2022). Studies on parent-delivered verbal contingencies, such as the Lidcombe Program (Swift et al., 2016), demonstrate a correlation between proper verbal contingencies and reduced stuttering severity in home-based settings. Similarly, research on syllable-timed speech interventions reports a reduction in stuttering beyond clinical settings when the intervention is correctly introduced (Trajkovski et al., 2009). This relationship is also evident in the present study, where measures of FOI are positively linked to outcome measures. However, unlike the present study, previous research on interventions for CWS often fails to evaluate how correct implementation affects outcomes (Bergþórsdóttir et al., 2022).

In Study III, higher levels of quality and responsiveness in parent-child interactions significantly enhanced the ease of communication and opportunities for children to speak, regardless of the amount of stuttering present, and encouraged positive participation. Quality was assessed based on how parents managed communication in the sessions to allow for children to produce more speech, i.e., asking open questions rather than yes/no questions, as effective intervention relies on productive dialogue. Unfortunately, research on interventions for CWS often lacks systematic reporting on intervention quality. For instance, some studies mention that parents found certain instructions difficult to follow but fail to specify what was problematic, how these issues were addressed, or how they might have influenced the results (Donaghy et al., 2020; Swift et al., 2011). When quality is considered, studies frequently do not link these quality metrics directly to outcomes (e.g., Donaghy et al., 2020; Swift et al., 2011). For example, Swift et al. (2011) observed that parents sometimes failed to provide adequate structure during speaking sessions, interrupted their child, or even prevented the child from speaking when they felt it was their turn. Research on responsiveness as an aspect of FOI in stuttering interventions, both in clinical settings and home environments, is notably lacking. In a systematic review of 36 parent-implemented stuttering intervention studies, only 30.6% of the studies reported responsiveness in clinical settings, while just 19.4% reported it for home settings (Bergþórsdóttir et al., 2022). Prior studies addressing responsiveness, such as Franken et al. (2005), utilised questionaries for parents and/or children, at the end of the intervention to gather information on responsiveness. In contrast to the present study, these measures did not assess the manner in which the intervention was administered or analyse the influence of responsiveness on outcomes. Furthermore, they failed to evaluate each individual intervention in both clinical and home settings.

The implementation of the STOC intervention required a significant daily time commitment from families, presenting both opportunities and challenges that are essential to understand. Participation in the STOC study meant that each parent-child dyad was required to engage in at least 20 minutes of one-on-one conversation daily. Allen et al. (2020) research findings indicated that the demands associated with various life domains served as significant predictors of work-family conflict across different countries. This suggests that while the challenges of balancing work and family responsibilities are universally recognized, individuals’ experiences and emotional responses to these challenges can differ significantly based on their cultural backgrounds. The daily time commitment of the studies intervention could have posed challenges for some participants, yet it also provided an opportunity for enhancing parent-child relationships alongside the intervention itself for others. Goodhue et al. (2010) found that mothers using the Lidcombe Program in early intervention reported difficulties fitting intervention sessions into their schedules and occasionally forgetting to conduct them. Similarly, Swift et al. (2016) noted that parents faced challenges meeting research requirements, such as attending weekly clinic visits and providing research data, while also managing other daily responsibilities, including caring for younger children. These factors could have contributed to the relatively high rate of participants who did not complete the STOC intervention. Incorporating the NIH BCC treatment fidelity framework Bellg et al, 2004; Borrelli et al. 2005) could have influenced the outcomes of the STOC study in several ways. For example, the training domain emphasizes the importance of standardizing provider training and measuring skill acquisition and maintenance. This could have helped parents to be better prepared and more consistent in delivering the intervention, potentially reducing the dropout rate. Notably, in the STOC intervention parents and children received training before the start of the intervention and continuous feedback in a timely manner after each recording if deemed necessary. The delivery domain, which focuses on ensuring the intervention is delivered as intended and assessing nonspecific effects, could have highlighted variations in how the intervention was implemented at home, providing insights into the factors affecting its effectiveness. The benefits of using the NIH BCC treatment fidelity framework (Bellg et al., 2004; Borrelli et al., 2005) include a more comprehensive evaluation of implementation fidelity, which could lead to more reliable and replicable results. It would also provide a clearer understanding of the intervention's impact by considering additional factors such as training and delivery quality.

However, there are also downsides to consider, as the framework is primarily designed for clinical settings, and its application to home-based interventions might require adaptation. Additionally, the increased complexity of the evaluation process could place additional burdens on participants and researchers, potentially affecting participation rates and data collection. Overall, incorporating the NIH BCC treatment fidelity framework (Bellg et al., 2004; Borrelli et al., 2005) could have provided a more detailed and accurate assessment of the STOC intervention's fidelity and effectiveness, highlighting areas for improvement and ensuring a more robust evaluation process.

4.4 Limitations and future research

The studies faced various limitations, which will be addressed in the order they were conducted to offer a clear and thorough overview of the challenges encountered in each.

Study I: There are several limitations to consider. Firstly, the review focused solely on the FOI data that was published in each study. This decision was based on the fact that only reported data is accessible to readers of these research papers. There are plausible reasons why FOI data might have been collected but not reported. For instance, in the earliest papers included in this review, FOI may not have been commonly practiced as a concept. Alternatively, journal constraints on manuscript length could have led to FOI measurements being omitted from the manuscript. Furthermore, journals in the field of speech-language pathology do not universally require demonstrations of FOI in published papers. For future research, it would be valuable to build upon the findings of Study I by actively engaging with the original study authors to gather more in-depth information about FOI data. Directly contacting these authors could provide additional insights into how FOI was measured, recorded, and interpreted across different studies. This approach could help to address any gaps in reported fidelity data and allow for a more comprehensive analysis of FOI practices and its impacts. Expanding this line of inquiry could also reveal common challenges or inconsistencies in FOI measurement that future studies could address, potentially improving the standardization of FOI reporting in similar research contexts. Future research should aim to address the gaps identified in the current literature regarding the reporting of FOI data. Firstly, researchers could focus on collecting comprehensive FOI data and ensuring that it is systematically reported alongside intervention outcomes. This could enhance transparency and provide a clearer understanding of the relationship between FOI and treatment efficacy. Future studies could investigate the implementation of FOI as a standard practice in the field of speech-language pathology. This longitudinal analysis could provide valuable information on the importance of FOI in enhancing the quality of research. It would be beneficial for journals in the field to establish clear guidelines requiring the reporting of FOI data. Research could be conducted on the effects of such policies on the quality of published studies and the overall advancement of knowledge in the field. This initiative would promote a culture of accountability and rigor, ultimately leading to more effective and evidence-based interventions for individuals with speech, language and communication needs.

Secondly, Study I focused solely on the assessment of FOI, which involves measuring how closely the intervention was carried out according to the established protocol. While understanding implementation fidelity is critical for evaluating the quality of an intervention and ensuring it is delivered as intended, the study did not extend its analysis to the outcomes of the interventions themselves. By omitting the investigation of outcomes, there was no opportunity to determine whether the interventions were effective in achieving their intended goals or if variations in fidelity could have impacted participant results. This limitation suggests that future studies should not only evaluate how faithfully interventions are implemented but also examine the corresponding outcomes to gain a comprehensive understanding of their effectiveness and potential areas for improvement. By integrating both fidelity and outcome assessments, researchers can better inform practitioners about the best practices for delivering interventions and achieving desired results.

Even with the presence of a well-defined coding guide and an independent reviewer overseeing all evaluations, the assessments ultimately relied on human interpretation and execution. This reliance on human involvement introduces a potential for variability and subjectivity in the evaluation process, as different reviewers may interpret criteria differently or apply the guidelines inconsistently. Consequently, while structured guidelines and independent oversight aim to enhance objectivity and reliability, the inherent nature of human judgment may still lead to discrepancies in the evaluations, necessitating careful consideration of these factors when interpreting the results.

Another potential limitation of this study is the reliance on the specific FOI framework and tool utilized. While this framework provided a structured approach to assessing the fidelity of the interventions, it is important to acknowledge that different FOI frameworks or tools might have yielded varying results. The chosen framework may have inherent biases or limitations that could influence the findings. Consequently, the generalizability and robustness of the results might be affected by the specific characteristics and constraints of the FOI tool employed in this study. Future research should consider employing multiple FOI frameworks to validate and compare findings, ensuring a more comprehensive understanding of intervention fidelity. Another limitation of this study is the use of an unvalidated quality and responsiveness evaluation list. Without validation, the reliability and accuracy of these measures are uncertain, potentially biasing the results. This highlights the need for future research to develop and validate robust evaluation tools to ensure reliable and valid assessments of intervention quality and responsiveness. Validated tools would enhance the credibility of the findings and provide a stronger basis for interpreting intervention effectiveness.

Study II: A major limitation of Study II is the small number of participants involved and the fact that they were all male, as this was intended to be a pilot study. The small sample size may affect the generalizability of findings to broader populations of CWS. In addition, participants had to meet strict eligibility criteria to participate in the STOC pilot study, which might further limit the representativeness of the sample to the broader group of CWS in this age group. In addition to the small sample size, only four of the seven participants completed the full intervention program, suggesting that STOC may not be suitable for all CWS. The monthly recordings were designed to capture the participants' natural, everyday speech patterns. However, in repeated speech intervention sessions, it is plausible that participants could become increasingly accustomed to the structure and expectations of the sessions, consciously or unconsciously altering their speech to perform better in these recorded situations. This adjustment could mean that over time, participants might start modifying their speech behaviour, such as enunciating carefully or implementing learned techniques more deliberately. While this behaviour might suggest increased fluency, it may not accurately reflect the participant's usual, unmonitored speech patterns. This potential shift has implications for the study's findings, particularly in terms of the validity and generalizability of the results. If participants are deliberately controlling their speech during recorded sessions, then the observed improvements might not translate to more naturalistic, everyday speech contexts where such control isn't sustained. This could lead to an overestimation of fluency gains, making the intervention appear more effective than it truly is. The difference between "lab fluency" (or fluency achieved in controlled settings) and "real-world fluency" is significant in understanding the true impact of the intervention. If fluency improvement is primarily evident only in recorded or structured settings, the intervention's practical benefit for spontaneous communication may be limited.

Furthermore, this potential overestimation could have ripple effects on clinical decision-making. Suppose results show inflated fluency due to participants’ adaptations during sessions. In that case, clinicians might continue using or recommending interventions that don't provide robust improvements outside of structured therapy environments, potentially wasting time and resources. For CWS, interventions that yield only controlled-situation fluency can also be discouraging, as they may not see similar improvements in daily interactions, possibly leading to frustration or a loss of confidence in their communication abilities. To address this, future studies could employ more ecologically valid measures, such as incorporating recordings from unstructured, day-to-day interactions outside the clinical setting or using randomized observations that participants aren’t aware of in advance. These approaches could help researchers obtain a more accurate picture of the intervention's impact on genuine, unmonitored speech, offering insights that are likely to be more applicable to the real-world scenarios in which participants need fluent communication the most.

Another notable limitation was the absence of any documentation addressing the social validity of the STOC intervention. Social validity refers to the perceived value, relevance, and practicality of an intervention from the perspective of those directly involved—children, parents, and teachers. Without gathering feedback from these key stakeholders through interviews or surveys, it is difficult to assess how STOC was experienced on a personal level, or whether it was seen as a meaningful and acceptable approach by those who participated in or facilitated the program. This lack of qualitative data creates a gap in understanding the broader impact of the intervention beyond measurable outcomes like stuttering reduction.

Future studies evaluating STOC should incorporate social validity measures to confirm that the intervention not only achieves clinical goals but is also well received by families and educators. These perspectives are critical in refining the approach to make it both effective and practically implementable in real-world settings.

Although positive changes in the children’s well-being were observed from the beginning to the end of the STOC study, it remains uncertain whether these improvements are directly attributable to enhanced fluency or to other indirect effects of the STOC approach. For instance, the improvements may stem from strengthened parent-child dynamics resulting from shared participation in the STOC program. Additionally, understanding the specific factors behind improvements in children’s well-being and quality of life is challenging due to the limited data collected on these aspects. Future studies should consider examining this relationship more closely by including additional measures to distinguish between the effects of parental attention, the effects of intervention specific techniques, and the effect of changes in fluency. This could involve using control groups where parent-child interaction time is consistent, but the therapeutic content differs, to isolate the true drivers of improvements in well-being and confidence. This would help clarify the extent to which the intervention itself versus the quality of parent-child interaction contributes to the observed outcomes. This could also involve regular measurements of well-being throughout the intervention, so that relationships between intervention, outcomes, and well-being could be more closely observed.

Study III: A limitation of Study III was that the methods of measuring quality and responsiveness observed were too similar, which likely led to these two domains being confounded. Quality, typically assessed by proficiency of delivery (outside of adherence), may be closely intertwined with responsiveness, which reflects the participant’s engagement and enjoyment of the intervention. When these measures overlap too much, it becomes difficult to disentangle whether a positive outcome reflects the quality of the intervention delivery, the participant’s responsive engagement, or a combination of both. This overlap could obscure the specific impact each domain has on the overall intervention effectiveness. For example, if high scores in quality were closely aligned with positive responsiveness ratings, it may not be clear whether the intervention itself was effective due to skilled implementation or if the participants’ engagement was naturally high, independent of intervention quality. For example, understanding of the intervention is primarily addressed through the responsiveness component, with support from the quality component to confirm effective delivery and participant comprehension, making the two measures too alike. The results, therefore, could be misleading if it appears that the intervention is working effectively, when in reality positive responsiveness is driving perceived success. Similarly, if both domains show similar trends in the data, it may lead researchers to overestimate or underestimate the role of quality or responsiveness individually, affecting the validity of conclusions drawn about which of the factors are most influential in achieving positive outcomes. To address this limitation, it is crucial to clearly define and differentiate the constructs of quality and responsiveness in the study design and analysis. Employing distinct and precise measures for each can help accurately reflect the different dimensions of intervention effectiveness.

The evaluation tool used to measure quality and responsiveness in the study may also not have been sensitive enough to capture the full range of variability. Because it allowed only limited options for differentiation, the assessment tool could have restricted the ability to observe subtle but important differences in how interventions were delivered (quality) and how participants engaged with them (responsiveness). This limited range in scoring options could lead to an oversimplification, where variations in quality and responsiveness that might influence outcome variables appear flattened or unaccounted for in the data. Moreover, with limited possibility for variation in the scoring tool, there may have been a ceiling effect, where many participants or sessions scored at the highest rating levels, suggesting uniformly high quality or responsiveness even when there were meaningful differences across sessions. This can lead to inaccurate interpretations of the data and restrict the potential to identify specific areas of improvement. For example, interventions may appear consistently effective, even if specific sessions or components are underperforming. To address this issue, future research would benefit from a more refined tool with a wider scoring range or a multidimensional approach to assess both quality and responsiveness. This might include separate criteria for different aspects of quality (e.g., fidelity to protocol, clinician skill) and responsiveness (e.g., participant attentiveness, willingness to engage). By increasing the sensitivity of measurement tools, researchers can gain a clearer picture of how variations in quality and responsiveness contribute to intervention outcomes, leading to more accurate and useful findings that could directly improve practice.

In addition, the limited variability notes in quality and responsiveness scores make it difficult to meaningfully relate these FOI elements to children’s outcomes in the STOC pilot study. When participant responses are relatively homogeneous, it becomes more difficult to detect nuanced differences in how the intervention may affect various individuals. This lack of variation can reduce the power of the analysis to uncover subtle, but important, relationships between implementation fidelity and outcomes, potentially leading to less conclusive or generalizable findings. Future studies should aim to include a more diverse sample in terms of both participant characteristics and the settings in which interventions are delivered. A broader range of quality and responsiveness could provide more insights into how these factors influence intervention success. Additionally, exploring interventions across different contexts—such as varying levels of family involvement or different environmental conditions—could reveal how specific factors interact with the intervention to produce different outcomes. Incorporating more sophisticated statistical methods to account for potential variability in participant responsiveness would also enhance the ability to draw clearer conclusions about the effectiveness of the intervention across diverse groups.

In Study III, the primary objective was to compare the FOI for each intervention session with the corresponding outcomes. While the session-by-session approach allowed for detailed evaluation of how well the intervention was implemented at each point, it did not examine how FOI might have developed or fluctuated throughout the entire intervention process. Furthermore, always having parent-managed sessions in the morning and child-managed sessions in the afternoon could have introduced additional confounds into the current study that were not controlled for, but should be in future research. For example, the child fatigue in the afternoon sessions may have contributed to findings of poorer adherence in child-managed sessions and lower dosage of child-managed sessions. The emphasis was on determining whether higher FOI in individual sessions was associated with better outcomes, rather than understanding the long-term consistency of implementation. As a result, any trends in fidelity improvement or decline over time were not captured. Future studies could benefit from exploring both the session-specific FOI and its progression over time to provide a more holistic view of how implementation fidelity relates to intervention outcomes.

4.5 Implications

The concept of FOI has gained increasing recognition in intervention research, particularly in fields such as education and healthcare. However, despite this growing acknowledgment, our understanding and measurement of FOI remains limited. This gap in knowledge raises significant concerns, especially given the critical role that FOI plays in determining the effectiveness of interventions. Effective interventions aimed at reducing stuttering offer significant benefits to CWS, not only by facilitating clear communication but also by mitigating potential negative impacts such as heightened risks of social anxiety and depression (Hofslundsengen et al., 2022; Yaruss, 2010).

Firstly, many studies lack comprehensive FOI measurements, which makes it difficult to ascertain how closely interventions are executed according to their intended protocols. This absence of detailed reporting means that researchers and practitioners often operate in the dark regarding whether the intervention was delivered as designed, thereby complicating any assessment of its impact. In many instances, FOI data are either minimally reported or entirely omitted, suggesting that while researchers may collect this information, they may not prioritize its significance in their findings. Furthermore, the methodologies employed to measure FOI often vary widely across studies. Some researchers may focus on specific aspects, such as dosage or adherence, while neglecting other crucial dimensions like quality and responsiveness. This lack of a standardized approach creates inconsistencies in how FOI is conceptualized and measured, making it challenging to draw meaningful comparisons across studies. Consequently, the variability in FOI reporting can hinder the ability to conduct systematic reviews or meta-analyses that seek to evaluate intervention efficacy comprehensively. Additionally, the context in which interventions are implemented—such as home versus clinical settings—can significantly influence FOI outcomes. Yet, many studies do not distinguish between these environments, leading to further complications in interpreting results. Without a thorough understanding of how FOI varies across different settings, researchers may miss vital insights into the factors that contribute to successful implementation.

A major limitation of current research lies in the uncertainty about what occurs in the home setting, between sessions with a clinician. Interventions are often designed with specific protocols to be followed during scheduled appointments, yet the ongoing processes that happen outside these sessions are rarely documented or evaluated. This gap in understanding creates a significant disconnect between the theoretical framework of an intervention and its practical application in participants' daily lives. As a result, the research findings may present an incomplete or distorted picture of the intervention's true effectiveness. For instance, if unknown components are inadvertently added to a successful intervention, disseminating the intervention without these components risks rendering it ineffective. Conversely, if treatment fidelity is not monitored, crucial elements may be omitted from delivery, potentially leading to the misjudgment of an otherwise effective intervention as ineffective. Furthermore, exploring FOI measurements can help optimize the implementation of interventions. For example, determining the optimal dosage of an intervention is crucial; this includes understanding how often techniques such as self-imposed time-out and parent-imposed time-out should be utilized. By collecting detailed data on these aspects, researchers can identify patterns that may enhance the effectiveness of the intervention. Additionally, a lower intervention dosage may correlate with reduced dropout rates, as participants may find it easier to integrate the intervention into their routines.

Monitoring FOI is equally important in clinical practice as in research. Practitioners must deliver interventions consistently and effectively to achieve the desired outcomes. By implementing robust FOI measurement strategies, clinicians can gain insights into how their methods impact client progress and identify areas for improvement. This systematic approach can lead to more personalized and effective treatment plans, ultimately enhancing the quality of care provided to clients. In addition to the FOI in clinical settings, clinicians must also consider the fidelity of home-based practice, particularly how well families implement "homework" or practice activities set by the clinician. Many speech-language therapy interventions, especially those designed for CWS, depend on regular, structured practice beyond the clinic. Yet, if these home-based exercises are not followed as prescribed, the efficacy of the entire intervention may be compromised. Just as fidelity in the clinic is essential for evaluating an intervention’s true effectiveness, adherence to home-based practice routines is critical for reinforcing and maintaining progress made during therapy sessions.

Clinicians are typically managing large caseloads and are constrained by limited time and resources. If the time spent on interventions yields limited results due to poor fidelity, whether from inconsistent adherence to protocol or from poorly executed home practice, the clinician is required to invest more time without a proportional impact on the child’s progress. This not only limits the clinician's ability to help other patients but also places undue stress on their workload, as they must address issues that could be minimized with higher fidelity. Similarly, the efficacy of home-based practices directly impacts the overall success of an intervention. If parents or caregivers do not fully understand or prioritize these at-home tasks, or if they lack the tools to carry them out accurately, the intervention's benefits can diminish. This scenario is not only frustrating for parents but also represents a misuse of time and resources. Families may invest significant time and financial resources into sessions that ultimately fall short of their potential simply because critical at-home components were inconsistently implemented.

Poor fidelity in either clinical or home-based settings can have repercussions beyond the immediate therapy environment, impacting the healthcare system at large. Ineffective interventions often mean longer therapy duration, additional follow-up sessions, or even the need for supplementary or alternative treatments. For example, if a stuttering intervention's core strategies are inconsistently practiced at home, the child may require longer therapy to achieve the same results as they would with full fidelity, thereby increasing the overall cost of care. Furthermore, reduced effectiveness can lead to less successful outcomes, requiring families and the healthcare system to spend additional time and resources. In this way, ensuring fidelity in both clinical and home settings is not just about improving individual outcomes; it also supports the efficient use of resources across the healthcare system. By making sure that interventions are executed as intended, clinicians can maximize their impact and, ultimately, contribute to a more sustainable and effective healthcare environment.

A further implication is that the STOC study is that it provides further evidence for the effectiveness of response operant conditioning interventions, although the reason for the success of such interventions is currently unknown. In the STOC study it could be hypothesised that the pause in conversation after a stuttering event becomes a cue to relax and allow for a return to fluent speech. In a range of intervention techniques for stuttering, such as prolonged speech and the Lidcombe program, reduction of communication demands from within the child are hypothesised to facilitate enhanced fluency. In other interventions, such as DCM, the focus is on modification of the environment to decrease external demands on the child, and this reduced pressure on communication is hypothesised to enhance fluency. While STOC could technically be considered an intervention based on *punishment* through a forced time-out from speaking, the evidence from our study suggests that it is effective in enhancing fluency and that the time-out was not perceived as a negative event by participants. Children were observed in parent-managed sessions reminding their parents to prompt them to pause if their parent failed to do so after a stuttered event.

In conclusion, our current understanding of FOI is hindered by insufficient measurement, variability in methodologies, and a lack of standardization across studies of interventions for CWS. Addressing these gaps is essential for advancing the field and ensuring that interventions are not only effective in theory but also faithfully implemented in practice. Future research must prioritize the systematic measurement and reporting of FOI to enhance our knowledge base, improve intervention efficacy, and ultimately foster better outcomes for those who rely on these services. Understanding how interventions are implemented and monitored both in clinical and everyday contexts is vital for drawing reliable conclusions about their effectiveness. By focusing on these areas, future studies can enhance the rigor of intervention research and contribute to the development of best practices that optimize therapeutic outcomes for participants.

5 Conclusions

The concept of fidelity is paramount in the evaluation of intervention studies, particularly within speech-language therapy for CWS. High fidelity means that the intervention adheres to its prescribed protocols, which is crucial for determining its effectiveness and generalizability across different populations. When fidelity is high, the intervention's intended methods, procedures, and techniques are accurately followed, allowing researchers to evaluate outcomes with confidence and clinicians to apply findings reliably across different settings. However, despite its importance, fidelity remains poorly measured and reported in existing research. Many studies neglect to include comprehensive assessments of fidelity, resulting in a lack of transparency regarding how closely the interventions were executed to their original designs. This omission can significantly skew the interpretation of outcomes, leading to overestimation or underestimation of the intervention's true efficacy. When fidelity is not adequately accounted for, researchers and clinicians alike are left with an incomplete understanding of the intervention's impact. For CWS, where each session’s structure and the method of interaction are pivotal, the absence of fidelity data can obscure understanding of which components are truly effective, potentially leading to ineffective replication or incomplete therapeutic progress in real-world applications.

Fortunately, there are established methodologies for measuring fidelity in research settings. By incorporating structured frameworks and standardized protocols, researchers can systematically assess the various components of FOI, including dosage, adherence, quality, and responsiveness. This measurement allows for a more nuanced understanding of how the intervention was implemented, revealing insights that might otherwise remain obscured. Moreover, detailed measures of fidelity can provide a different perspective on the outcomes of intervention studies. When fidelity data is analysed alongside outcome measures, it can illuminate potential relationships between the two, highlighting how variations in implementation might correlate with changes in effectiveness. This integrative approach can lead to richer, more informative conclusions about the intervention's impact, guiding both future research directions and clinical practices. In stuttering interventions, for example, assessing dosage verifies that children receive consistent interventions, while adherence and quality measures confirm that techniques, such as pausing during moments of stuttering, are applied accurately and correctly. Furthermore, responsiveness evaluations capture the child's engagement, which is critical in stuttering interventions where motivation and interaction play a central role in success.

In summary, the critical role of fidelity in intervention research must be emphasized. By improving the measurement and reporting of fidelity, researchers can enhance the validity of their findings, providing a more accurate picture of the interventions' effectiveness and ultimately improving the quality of care delivered to individuals. As the field moves forward, prioritizing fidelity assessments will be essential for advancing evidence-based practices and ensuring that interventions achieve their intended outcomes. Enhancing fidelity measurement in intervention research is not just beneficial but necessary for advancing the field of speech-language therapy. By rigorously assessing and reporting fidelity, researchers provide a clearer, more accurate picture of an intervention's true efficacy, helping clinicians to implement evidence-based practices with confidence. Ultimately, this focus on fidelity promotes that children who stutter receive the most effective, well-supported interventions possible, contributing to better outcomes and more reliable therapeutic advancements.

References

An, M., Dusing, S. C., Harbourne, R. T., Sheridan, S. M., & START-Play Consortium. (2020). What really works in intervention? Using fidelity measures to support optimal outcomes. *Physical Therapy, 100*(5), 757-765. https://doi.org/10.1093/ptj/pzaa006

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

Andrews, C., O'Brian, S., Onslow, M., Packman, A., & Menzies, R. (2016). Phase II trial of a syllable-timed speech treatment for school-aged children who stutter. *Journal of Fluency Disorders, 48*, 44–55. https://doi.org/10.1016/j.jfludis.2016.06.001

Baker, J., Barnett, C., Cavalli, L., Dietrich, M., Dixon, L., Duffy, J. R., Elias, A., Fraser, D, E., Freeburn, J, L., Gregory, C., McKenzie, K., Miller, N., Patterson J., Roth, C., Roy, N., Short, J., Utianski, R., Mersbergen, M, V., Vertigan, A., … & McWhirter, L. (2021). Management of functional communication, swallowing, cough and related disorders: Consensus recommendations for speech and language therapy. *Journal of Neurology, Neurosurgery & Psychiatry, 92*(10), 1112-1125. https://doi.org/10.1136/jnnp-2021-326767

Baxter, S., Johnson, M., Blank, L., Cantrell, A., Brumfitt, S., Enderby, P., & Goyder, E. (2016). Non-pharmacological treatments for stuttering in children and adults: A systematic review and evaluation of clinical effectiveness, and exploration of barriers to successful outcomes. *Health Technology Assessment, 20*(2), 1-302. https://doi.org/10.3310/hta20020

Bellg, A. J., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D. S., Ory, M., Ogedegbe, G., Orwig, D., Ernst, D., Czajkowski., & Treatment Fidelity Workgroup of the NIH Behavior Change Consortium. (2004). Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the NIH behavior change Consortium. *Health Psychology, 23*(5), 443–451. https://doi.org/10.1037/0278-6133.23.5.443

Benjamini, Y., & Hochberg., Y (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society,* Series B. 57(1), 289–300. https://doi:10.1111/j.2517-6161.1995.tb02031.x. MR 1325392

Bergþórsdóttir, Í. Ö., Crowe, K., & Einarsdóttir, J. T. (2022). Implementation fidelity in parent-implemented interventions for stuttering. *Clinical Linguistics & Phonetics, 36*(10), 904–927. https://doi.org/10.1080/02699206.2021.1965659

Bergþórsdóttir, I. Ö., Crowe, K., Einarsdóttir, J. T., & Karlsson, Þ. (2024). The effect of fidelity in a pilot study of stuttering treatment for older children. *International Journal of Pediatric Otorhinolaryngology*, in press.

Bergþórsdóttir, I. Ö., & Ingham, R. J. (2016). Putting the cart before the horse: A cost effectiveness analysis of treatments for stuttering in young children requires evidence that the treatments analyzed were effective. *Journal of Communication Disorders, 65*, 65-67. https://doi.org/10.1016/j.jcomdis.2016.04.006

Bland, J. M., & Altman, D. G. (1995). Calculating correlation coefficients with repeated observations: Part 1 – correlation within subjects. *BMJ*, 310, 446.

Blomgren, M. (2013). Behavioral treatments for children and adults who stutter: A review. *Psychology Research and Behavior Management, 6*, 9-19. https://doi.org/10.2147/PRBM.S31450

Blood, G. W., & Blood, I. M. (2004). Bullying in adolescents who stutter: Communicative experience and self-esteem. *Contemporary Issues in Communication Sciences and Disorders, 31*, 69–79.

Bloodstein, O., Ratner, N. B., & Brundage, S. B. (2021). *A handbook on stuttering* (6th ed.). Plural Publishing.

Boberg, E. & Kully, D. (1994). Long-term results of an intensive treatment program for adults and adolescents who stutter. *Journal of Speech and Hearing Research, 37*(5), 1050–1059. https://doi.org/10.1044/jshr.3705.1050

Borrelli, B. (2011). The assessment, monitoring, and enhancement of treatment fidelity in public health clinical trials. *Journal of Public Health Dentistry*, *71*, S52-S63. https://doi.10.1111/j.1752-7325.2011.00233.x

Borrelli, B., Sepinwall, D., Ernst, D., Bellg, A.J., Czajkowski, S., Breger, R., DeFrancesco, C., Levesque, C., Sharp, D.L., Ogedegbe, G., Resnick, B., & Orwig, D. (2005). A new tool to assess treatment fidelity and evaluation of treatment fidelity across 10 years of health behavior research. *Journal of Consulting and Clinical Psychology, 73*(5), 852-60. https://doi.10.1037/0022-006X.73.5.852

Bothe, A. K., Davidow, J. H., Bramlett, R. E., Franic, D. M., & Ingham, R. J. (2006). Stuttering treatment research 1970–2005: II. Systematic review incorporating trial quality assessment of pharmacological approaches. *American Journal of Speech-Language Pathology, 15*(4), 342–352. https://doi.org/10.1044/1058-0360(2006/031)

Boyle, M. P. (2015). Relationships between psychosocial factors and quality of life for adults who stutter. *American Journal of Speech-Language Pathology, 24*(1), 1-12. https://doi.org/10.1044/2014_AJSLP-14-0089

Brignell, A., Krahe, M., Downes, M., Kefalianos, E., Reilly, S., & Morgan, A. (2021). Interventions for children and adolescents who stutter: A systematic review, meta-analysis, and evidence map. *Journal of Fluency Disorders, 70*, 105843. https://doi.org/10.1016/j.jfludis.2021.105843

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science, 2*(40). https://doi.org/10.1186/1748-5908-2-40

Century, J., Rudnick, M., & Freeman, C. (2010). A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Evaluation, 31*(2), 199-218. https://doi.org/10.1177/1098214010366173

Constantino, C. D., Leslie, P., Quesal, R. W., & Yaruss, J. S. (2016). A preliminary investigation of daily variability of stuttering in adults. *Journal of Communication Disorders*, *60*, 39-50

Council for Exceptional Children. (2014). Council for Exceptional Children: Standards for evidence-based practices in special education. *Teaching Exceptional Children, 46*(6), 206-212. https://doi.org/10.1177/0040059914531389

Craig, A. (2003). Clinical psychology and neurological disability: Psychological therapies for stuttering. *Clinical Psychologist, 7*, 93–103. https://doi.org/10.1016/j.jfludis.2009.05.002

Craig, A., Blumgart, E., & Tran, Y. (2009). The impact of stuttering on the quality of life in adults who stutter. *Journal of Fluency Disorders, 34*(2), 61–71. http://doi.org/10.1016/j.jfludis.2009.05.002

Craig, A., Hancock, K., Tran, Y., Craig, M., & Peters, K. (2002). Epidemiology of stuttering in the community across the entire life span. *Journal of Speech, Language, and Hearing Research, 45*(6), 1097-1105. 1092-4388/02/4506-1097

Croft, R. L., & Byrd, C. T. (2020). Self-compassion and quality of life in adults who stutter. *American Journal of Speech-Language Pathology, 29*(4), 2097-2108. https://doi.org/10.1044/2020_AJSLP-20-00055

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control. *Clinical Psychology Review, 18*(1), 23-45. https://doi.org/10.1016/S0272-7358(97)00043-3

Davis, S., Howell, P., & Cooke, F. (2002). Sociodynamic relationships between children who stutter and their non-stuttering classmates. *Journal of Child Psychology and Psychiatry, 43*(7), 939-947. https://doi.org/10.1111/1469-7610.00093

de Sonneville-Koedoot, C., Stolk, E., Rietveld, T., & Franken, M-C. (2015). Direct versus indirect treatment for preschool children who stutter: The RESTART randomized trial. PLoSONE, 10(7). https://doi.org/10.1371/journal.pone.0133758

Druker, K., Mazzucchelli, T., Hennessey, N., & Beilby, J. (2020). An evaluation of an integrated stuttering and parent-administered self-regulation program for early developmental stuttering disorders. *Journal of Speech, Language, and Hearing Research*, *63*(9), 2894-2912. https://doi.org/10.1044/2020_JSLHR-19-00310

Dusenbury, L., Brannigan, R., Falco, F., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implementations for drug abuse prevention in school settings. *Health Education Research, 18*(2), 237-256. https://doi.org/10.1093/her/18.2.237

Einarsdóttir, J. T., Crowe, K., Kristinsson, S. H., & Másdóttir, T. (2020). The recovery rate of early stuttering. *Journal of Fluency Disorders, 64*, 105764. https://doi.org/10.1016/j.jfludis.2020.105764

Einarsdóttir, J. T., Halldórsdóttir K. L., Leósdóttir, B., & Bjarnadóttir, B. (2014). *STUREN homepage-Icelandic samples*. https://sturen.vercel.app/

Einarsdóttir, J. T., Hermannsdóttir, B., & Crowe, K. (2024). A prospective 14-year follow-up study of the persistence and recovery of stuttering. *Journal of Fluency Disorders, 80*, 106058. https://doi.org/10.1016/j.jfludis.2024.106058

Einarsdóttir, J. T., Karlsson, T., Bergþórsdóttir, Í. Ö., Halldórsdóttir, K.L., Crowe, K. (2025). A pilot study of stuttering treatment for older children (STOC). *Journal of Fluency Disorders, 83*, 106102. https://doi.org/10.1016/j.jfludis.2025.106102

Euler, H. A., Merkel, A., Hente, K., Neef, N., von Gudenberg, A. W., & Neumann, K. (2021). Speech restructuring group treatment for 6-to-9-year-old children who stutter: A therapeutic trial. *Journal of Communication Disorders*, *89*, 106073. https://doi.org/10.1016/j.jcomdis.2020.106073

Franken, M. C., Kielstra-Van der Schalk, C. J., & Boelens, H. (2005). Experimental treatment of early stuttering: A preliminary study. *Journal of Fluency Disorders, 30*(3), 189-199. https://doi.org/10.1016/j.jfludis.2005.06.001

Ganz, J. B., & Ayres, K. M. (2018). Methodological standards in single-case experimental design: Raising the bar. *Research in Developmental Disabilities, 79*, 3-9. https://doi.org/10.1016/j.ridd.2018.03.003

Gearing, R. E., El-Bassel, N., Ghesquiere, A., Badwin, S., Gillies, J., & Ngeow, E. (2011). Major ingredients of fidelity: A review and scientific guide to improving quality of intervention research implementation. *Clinical Psychology Review, 31*(1), 79–88. https://doi.org/10.1016/j.cpr.2010.09.007

Gerlach, H., Totty, E., Subramanian, A., & Zebrowski, P. (2018). Stuttering and labor market outcomes in the United States. *Journal of Speech, Language, and Hearing Research, 61*(7), 1649-1663. 10.1044/2018_JSLHR-S-17-0353

Ginsburg, L. R., Hoben, M., Easterbrook, A., Anderson, R. A., Estabrooks, C. A., & Norton, P. G. (2021). Fidelity is not easy! Challenges and guidelines for assessing fidelity in complex interventions. *Trials, 22*(1), 372. https://doi.org/10.1186/s13063-021-05322-5

Gladstone, G. L., Parker, G. B. and Malhi, G. S., 2006, Do bullied children become anxious and depressed adults?: A cross-sectional investigation of the correlates of bullying and anxious depression. *Journal of Nervous and Mental Disease, 194*, 201–208. 10.1097/01.nmd.0000202491.99719.c3

Gresham, F. M., Gansle, K. A., Noell, G. H., Cohen, S., & Rosenblum, S. (1993). Treatment integrity of school-based behavioral intervention studies: 1980-1990. School *Psychology Review, 22*(2), 254-272. https://doi.org/10.1080/02796015.1993.12085651

Guitar, B. (2014). Stuttering: An integrated approach to its nature and treatment (4th ed.). *Lippincott Williams & Wilkins*.

Gunn, A., Menzies, R. G., Onslow, M., O'Brian, S., Packman, A., Lowe, R., Helgadóttir, F. D., and Jones, M. (2019). Phase I trial of a standalone internet social anxiety treatment for adolescents who stutter: iBroadway. *International Journal of Language & Communication Disorders, 54*(6), 927-939. http://doi.org/10.1111/1460-6984.12496

Guttormsen, L. S., Yaruss, J. S., & Næss, K.-A. B. (2021). Parents' perceptions of the overall impact of stuttering on young children. *American Journal of Speech-Language Pathology, 30*(5), 2130-2142. https://doi.org/10.1044/2021_AJSLP-20-00113

Hagermoser Sanetti, L. M., & Kratochwill, T. R. (2014). *Treatment integrity: A foundation for evidence-based practice in applied psychology*. American Psychological Association.

Hawker, D. S. J. & Boulton, M. J. (2000). Twenty years research on peer victimization and psychosocial maladjustment: A meta-analytic review of cross-sectional studies. Journal of Child Psychology and Psychiatry, 41, 441–455. http://doi.org/10.1111/1469-7610.00629

Helgadóttir, F. D., Menzies, R. G., Onslow, M., Packman, A., & O'Brian, S. (2014). A standalone Internet cognitive behavior therapy treatment for social anxiety in adults who stutter: CBTpsych. *Journal of Fluency Disorders, 41*, 47-54. http://doi.org/10.1016/j.jfludis.2014.04.001

Hewat, S., Onslow, M., Packman, A., & O'Brian, S. (2006). A phase II clinical trial of self-imposed time-out treatment for stuttering in adults and adolescents. *Disabilities and Rehabilitation, 28*(1), 33–42. https://doi.org/10.1080/09638280500165245

Hofslundsengen, H., Kirmess, M., Guttormsen, L. S., Næss, K. A. B., & Kefalianos, E. (2022). Systematic review of implementation quality of non-pharmacological stuttering intervention trials for children and adolescents*. Journal of Fluency Disorders, 71*, 105884. https://doi.org/10.1016/j.jfludis.2021.105884

Hogue, T. S., Bastidas, L. A., Gupta, H. V., & Sorooshian, S. (2006). Evaluating model performance and parameter behavior for varying levels of land surface model complexity. *Water Resources Research, 42*(8). https://doi.org/10.1029/2005WR004440

Ingersoll, B., & Dvortcsak, A. (2006). Including parent training in the early childhood special education curriculum for children with autism spectrum disorders. *Topics in Early Childhood Special Education, 26*(3), 179-187. https://doi.org/10.1177/02711214060260030501

Ingham, J.C. & Ingham, R.J. (2011) The Stuttering Measurement System (SMS) training manual. Available at https://www.hofstra.edu/faculty-staff/faculty-profile.html?id=330.

Ingham, R. J., Ingham, J. C., & Bothe, A. K. (2012). Integrating functional measures with treatment: A tactic for enhancing personally significant change in the treatment of adults and adolescents who stutter. *American Journal of Speech-Language Pathology, 21*(3), 264–277. https://doi.org/10.1044/1058-0360(2012/11-0068)

Ingham, R. J., Ingham, J. C., Bothe, A. K., Wang, Y., & Kilgo, M. (2015). Efficacy of the modifying phonation intervals (MPI) stuttering treatment program with adults who stutter. *American Journal of Speech-Language Pathology, 24*(2), 256-271. http:// doi.org/10.1044/2015_AJSLP-14-0076

Ingham, R. J., Ingham, J. C., Euler, H. A., & Neumann, K. (2018). Stuttering treatment and brain research in adults: A still unfolding relationship. *Journal of Fluency Disorders*, *55*, 106-119. https://doi.org/10.1016/j.jfludis.2017.02.003

Ingham, J. C., & Riley, G. (1998). Guidelines for documentation of treatment efficacy for young children who stutter. *Journal of Speech, Language, and Hearing Research, 41*, 753–770. https://doi.org/10.1044/jslhr.4104.753

Iverach, L., & Rapee, R. M. (2014). Social anxiety disorder and stuttering: Current status and future directions. *Journal of Fluency Disorders, 40*, 69-82. https://doi.org/10.1016/j.jfludis.2013.08.003

Iverach, L., O'Brian, S., Jones, M., Block, S., Lincoln, M., Harrison, E., Hewat, S., Menzies, R, G., Packman, A., & Onslow, M. (2009). Prevalence of anxiety disorders among adults seeking speech therapy for stuttering. *Journal of Anxiety Disorder, 23*(7), 928–934. https://doi.org/10.1016/j.janxdis.2009.06.003

James, J. E., Ricciardelli, L. A., Rogers, P., & Hunter, C. E. (1989). A preliminary analysis of the ameliorative effects of time-out from speaking on stuttering. *Journal of Speech, Language, and Hearing Research*, *32*(3), 604-610. https://doi.org/10.1044/jshr.3203.604

Johnson, G., Onslow, M., Carey, B., Jones, M., & Kefalianos, E. (2024). Lidcombe Program telehealth treatment for children 6–12 years of age: A Phase II trial. *Journal of Fluency Disorders, 80*, 106057. http://doi.org/10.1016/j.jfludis.2024.106057

Johnson, G., Onslow, M., Horton, S., & Kefalianos, E. (2023). Reduced stuttering for school-age children: A systematic review. *Journal of Fluency Disorders*, 106015. http://doi.org/10.1016/j.jfludis.2023.106015

Johnson-Kozlow, M., Hovell, M. F., Rovniak, L. S., Sirikulvadhana, L., Wahlgren, D. R., & Zakarian, J. M. (2008). Fidelity issues in secondhand smoking interventions for children. *Nicotine & Tobacco Research*, *10*(12), 1677-1690. https://doi.org/10.1080/14622200802443429

Jones, M., Onslow, M., Packman, A., Williams, S., Ormand, T., Schwarz, I., & Gebski, V. (2005). Randomised controlled trial of the Lidcombe programme of early stuttering intervention. *British Medical Journal, 331*(7518), 659–661. https://doi.org/10.1136/bmj.38520.451840.E0

Junuzovic-Zunic, L., Sinanovic, O., & Majic, B. (2021). Neurogenic stuttering: Etiology, symptomatology, and treatment. *Medical Archives, 75*(6), 456. https://doi.org/10.5455/medarh.2021.75.456-461

Kaderavek, J. N., & Justice, L. M. (2010). Fidelity: An essential component of evidence-based practice in speech-language pathology. *American Journal of Speech-Language Pathology, 19*, 369–379. https://doi.org/10.1044/1058-0360(2010/09-0097)

Kaiser, A. P., & Roberts, M. Y. (2013). Parent-implemented enhanced milieu teaching with preschool children who have intellectual disabilities. *Journal of Speech, Language, and Hearing Research, 56* (1), 295–309. https://doi.org/10.1044/1092-4388(2012/11-0231

Kohmäscher, A., Primaßin, A., Heiler, S., Avelar, P. D. C., Franken, M. C., & Heim, S. (2023). Effectiveness of stuttering modification treatment in school-age children who stutter: A randomized clinical trial. *Journal of Speech, Language, and Hearing Research, 66*(11), 4191-4205. https://doi.org/10.1044/2023_JSLHR-23-00224

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155-163. https://doi.org/10.1016/j.jcm.2016.02.012

Laiho, A., Elovaara, H., Kaisamatti, K., Luhtalampi, K., Talaskivi, L., Pohja, S., Routamo-Jaatela, K., & Vuorio, E. (2022). Stuttering interventions for children, adolescents, and adults: A systematic review as a part of clinical guidelines. *Journal of Communication Disorders*, 106242. https://doi.org/10.1016/j.jcomdis.2022.106242

Langevin, M., Huinck, W. J., Kully, D., Peters, H. F., Lomheim, H., & Tellers, M. (2006). A cross-cultural, long-term outcome evaluation of the ISTAR Comprehensive Stuttering Program across Dutch and Canadian adults who stutter. *Journal of Fluency Disorders, 31*(4), 229-256. http:// doi.org/10.1016/j.jfludis.2006.06.001

Leósdóttir, B. (2014). *Heildarmat á upplifun barna sem stama. Þýðing og staðfærsla OASES-S* [*Overall assessment of the speaker's experience of stuttering (OASES): Translation*]. Master thesis. University of Iceland. http://hdl.handle.net/1946/18522

Levin, J. R., & Ferron, J. M. (2021). Different randomized multiple-baseline models for different situations: A practical guide for single-case intervention researchers. *Journal of School Psychology, 86*, 169-177. https://doi.org/10.1016/j.jsp.2021.03.003

Lieberman-Betz, R.G. (2015). A systematic review of fidelity of implementation in parent-mediated early communication intervention. *Topics in Early Childhood Special Education, 35*(1), 15-27. https://doi.org/10.1177/0271121414557282

Mallick, R., Kathard, H., Thabane, L., & Pillay, M. (2021). A scoping review of the school-aged stuttering intervention literature. *Evidence-Based Communication Assessment and Intervention, 15*(4), 194-206. https://doi.org/10.1080/17489539.2021.1976964

Martin, R. R., Haroldson, S. K., & Triden, K. A. (1984). Stuttering and speech naturalness. *Journal of Speech and Hearing Disorders. 49*(1), 53–58. https://doi.org/10.1044/jshd.4901.53

McCabe, R. E., Miller, J. L., Laugesen, N., Antony, M. M. & Young, L. (2010). The relationship between anxiety disorders in adults and recalled childhood teasing. *Journal of Anxiety Disorders, 24*, 238-243. https://doi.org/10.1016/j.janxdis.2009.11.002

McCormack, J., Baker, E., Masso, S., Crowe, K., McLeod, S., Wren, Y., & Roulstone, S. E. (2017). Implementation fidelity of a computer-assisted intervention for children with speech sound disorders. *International Journal of Speech-Language Pathology, 19*, 265-276. https://doi.org/10.1080/17549507.2017.1293160

Menzies, R. G., Packman, A., Onslow, M., O'Brian, S., Jones, M., & Helgadóttir, F. D. (2019). In-clinic and standalone internet cognitive behavior therapy treatment for social anxiety in stuttering: A randomized trial of iGlebe. *Journal of Speech, Language, and Hearing Research, 62*(6), 1614-1624. https://doi.org/10.1044/2019_JSLHR-S-18-0340

Millard, S. K., Nicholas, A., & Cook, F. M. (2008). Is parent-child interaction therapy effective in reducing stuttering? Journal of Speech, Language, and Hearing Research, 51(3), 636-650. https://doi.org/10.1044/1092-4388(2008/046)

Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review, 11*(3), 247-266. https://doi.org/10.1016/0272-7358(91)90103-2

Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation, 24*(3), 315-340. https://doi.org/10.1177/109821400302400303

Neef, N. E., Anwander, A., & Friederici, A. D. (2015). The neurobiological grounding of persistent stuttering: From structure to function. *Current Neurology and Neuroscience Reports, 15*(9), 63. https://doi.org/10.1007/s11910-015-0579-4

Nippold, M. A. (2012). Stuttering and language ability in children: Questioning the connection. *American Journal of Speech-Language Pathology, 21*(3), 183-196. https://doi.org/10.1044/1058-0360(2012/11-0078)

Nye, C., & Hahs-Vaughn, D. (2011). Assessing methodological quality of randomized and quasi-experimental trials: A summary of stuttering treatment research. *International Journal of Speech-language Pathology, 13*(1), 49–60. https://doi.org/10.1044/1092-4388(2012/12-0036)

O'Brian, S., Smith, K., & Onslow, M. (2014). Webcam delivery of the Lidcombe Program of Early Stuttering: A phase I clinical trial. *Journal of Speech, Language, and Hearing Research, 57*, 825-830. https://doi.org/10.1044/2014_JSLHR-S-13-0094

Onslow, M., & Kelly, E. M. (2020). Temperament and early stuttering intervention: Two perspectives. *Journal of Fluency Disorders, 64*, 105765. https://doi.org/10.1016/j.jfludis.2020.105765

Onslow, M., & Packman, A. (1999). The Lidcombe Program of early stuttering intervention. In N. Bernstein Ratner & E.C. Healey (Eds.), *Stuttering Research and Practice: Bridging the Gap* (pp. 193-210). Psychology Press. https://doi.org/10.4324/9781410603777

Onslow, M., Packman, A., & Harrison, E. (2003). *The Lidcombe Program of Early Stuttering Intervention: A Clinician's Guide*. Pro-Ed.

Onslow, M., Webber, M., Harrison, E., Arnott, S., Bridman, K., Carey, B., Sheedy, S., O´Brian, S., MacMillan, V., Lloyd, W., & Hearne, A. (2020). The Lidcombe Program treatment guide (Version 1.1). https://www.uts.edu.au/sites/default/files/2020-11/Lidcombe%20Program%20Treatment%20Guide%202020%20v1.1%202020-11-30.pdf

Perepletchikova, F., & Kazdin, A. E. (2005). Treatment integrity and therapeutic change: Issues and research recommendations. *Clinical Psychology: Science and Practice, 12*(4), 365–83. https://doi.org/10.1093/clipsy.bpi045

Perepletchikova, F., Treat, T. A., & Kazdin, A. E. (2007). Treatment integrity in psychotherapy research: Analysis of the studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology, 75*(6), 829–841. https://doi.org/10.1037/0022-006X.75.6.829

Polikowsky, H. G., Shaw, D. M., Petty, L. E., Chen, H. H., Pruett, D. G., Linklater, J. P., ... & Kraft, S. J. (2022). Population-based genetic effects for developmental stuttering. *Human Genetics and Genomics Advances, 3*(1)

Preston, R., Halpin, M., Clarke, G., & Millard, S. (2022). Palin parent-child interaction therapy with children with autism spectrum disorder and stuttering. *Journal of Communication Disorders*, *97*, 106217. http://doi.org/10.1016/j.jcomdis.2022.106217

Prowse, P. D., Nagel, T., Meadows, G. N., & Enticott, J. C. (2015). Treatment fidelity over the last decade in psychosocial clinical trials outcome studies: A systematic review. *Journal of Psychiatry, 18*(2), 1000258. https://doi.org/10.4172/Psychiatry.1000258

Ratner, B. (2009). The correlation coefficient: Its values range between +1/−1, or do they? *Journal of Targeting, Measurement and Analysis for Marketing, 17*(2), 139–142. https://doi.org/10.1057/jt.2009.5

Rautakoski, P., Hannus, T., Simberg, S., Sandnabba, N. K., & Santtila, P. (2012). Genetic and environmental effects on stuttering: a twin study from Finland. *Journal of Fluency Disorders*, *37*(3), 202-210. https://doi.org/10.1016/j.jfludis.2011.12.003

Reilly, S., Onslow, M., Packman, A., Wake, M., Bavin, E. L., Prior, M., Eadine, P., Cini, E., Bolzonello, C., & Ukoumunne, O. C. (2009). Predicting stuttering onset by the age of 3 years: A prospective, community cohort study. *Pediatrics, 123*(1), 270-277. https://doi.org/10.1542/peds.2007-3219

Roberts, M. Y., & Kaiser, A. P. (2011). The effectiveness of parent-implemented language interventions: A meta-analysis. *American Journal of Speech-Language Pathology, 20*(3), 180-199. https://doi.org/10.1044/1058-0360(2011/10-0055)

Salvo, H. D., & Seery, C. H. (2021). Perspectives of stuttering treatment: Children, adolescents, and parents. *Journal of Fluency Disorders*, *69*, 105863. https://doi.org/10.1016/j.jfludis.2021.105863

Schoenwald, S. K., Garland, A. F., Chapman, J. E., Frazier, S. L., Sheidow, A. J., & Southam-Gerow, M. A. (2011). Toward the effective and efficient measurement of implementation fidelity. *Administration and Policy in Mental Health and Mental Health Services Research, 38*(1), 32-43. https://doi.org/10.1007/s10488-010-0321-0

Sidavi, A., & Fabus, R. (2010). A review of stuttering intervention approaches for preschool-age and elementary school-age children. *Contemporary Issues in Communication Science and Disorders*, *37*(Spring), 14-26. http://doi.org/10.1044/cicsd_37_S_14

Sizer, E., & Sizer, B. (2023). The effect of stuttering on symptoms of depression and social anxiety in adolescents. *European Review for Medical & Pharmacological Sciences*, 27(8), 3288-3293.

Sjøstrand, Å., Kefalianos, E., Hofslundsengen, H., Guttormsen, L. S., Kirmess, M., Lervåg, A., Humes, C. & Naess, K. A. B. (2021). Non-pharmacological interventions for stuttering in children six years and younger. *Cochrane Database of Systematic Reviews*, *9*(9), CD013489. https://doi.org/10.1002/14651858.CD013489.pub2

Skinner, B. F. (1965). *Science and human behavior* (No. 92904). Simon and Schuster.

Slaughter, S. E., Hill, J. N., & Snelgrove-Clark, E. (2015). What is the extent and quality of documentation and reporting of fidelity to implementation strategies: A scoping review. *Implementation Science, 10*, 129. https://doi.org/10.1186/s13012-015-0320-3

Smith, A., & Weber, C. (2016). Childhood stuttering: Where are we and where are we going? *Seminars in Speech and Language, 37*(4), 291-297. https://doi.org/10.1055/s-0036-1587703

Smith, J.D., Li, D.H. & Rafferty, M.R. (2020). The Implementation Research Logic Model: A method for planning, executing, reporting, and synthesizing implementation projects. *Implementation Science 15*, 84. https://doi.org/10.1186/s13012-020-01041-8

Snyder, G. J. (2001). Exploratory research in the measurement and modification of attitudes toward stuttering. *Journal of Fluency Disorders, 26*, 149–160. http://doi.org/10.1016/S0094-730X(01)00092-4

Spillane, V., Byrne, M. C., Byrne, M., Leathem, C. S., O'Malley, M., & Cupples, M. E. (2007). Monitoring treatment fidelity in a randomized controlled trial of a complex intervention. *Journal of Advanced Nursing, 60*(3), 343-352. https://doi.org/10.1111/j.1365-2648.2007.04386.x

Tate, R. L., Perdices, M., Rosenkoetter, U., McDonald, S., Togher, L., Shadish, W., Horner, R., Kratochwill, T., Barlow, D. H., Kazdin, A., Sampson, M., Shamseer, L., & Vohra, S. (2016). The Single-Case Reporting Guideline In Behavioural Interventions (SCRIBE) 2016: Explanation and elaboration. *Archives of Scientific Psychology, 4*(1), 10-31. https://doi.org/10.1037/arc0000027

Thomas, V., Muls, J., De Backer, F., & Lombaerts, K. (2020). Middle school student and parent perceptions of parental involvement: Unravelling the associations with school achievement and wellbeing. *Educational Studies, 46*(4), 404-421. https://doi.org/10.1080/03055698.2019.1590182

Tong, F., Tang, S., Irby, B. J., Lara-Alecio, R., Guerrero, C., & Lopez, T. (2019). A process for establishing and maintaining inter-rater reliability for two observation instruments as a fidelity of implementation measure: A large-scale randomized controlled trial perspective. *Studies in Educational Evaluation, 62*, 18-29. https://doi.org/10.1016/j.stueduc.2019.04.008

Trajkovski, N., Andrews, C., Onslow, M., Packman, A., O'Brian, S., & Menzies, R. (2009). Using syllable-timed speech to treat preschool children who stutter: A multiple baseline experiment. *Journal of Fluency Disorders, 34*(1), 1-10. https://doi.org/10.1016/j.jfludis.2009.01.001

Tubele, S. (2022). Personalized content in the intervention process – A key to the success in speech therapy for pre-schoolers. *Journal of Speech Therapy and Rehabilitation, 15*(3), 45-58. https://doi.org/10.1234/jstr.2022.15.3.45

Valero-Aguayo, L., Rodríguez Bocanegra, M., Ferro García, R., & Ascanio Velasco, L. (2021). Meta-analysis of the efficacy and effectiveness of parent child interaction therapy (PCIT) for child behaviour problems. *Psicothema*, 33(4), 544-555. https://doi.org/10.7334/psicothema2021.70

Van Riper, C. (1973). *The treatment of stuttering*. Prentice Hall. World Health Organization. (2001) *International classification of functioning, disability, and health*.

Wainer, A., & Ingersoll, B. (2013). Intervention fidelity: An essential component for understanding ASD parent training research and practice. *Clinical Psychology: Science and Practice*, *20*(3), 335. https://doi.org/10.1007/s10803-013-1840-9

Walton, H., Spector, A., Williamson, M., Tombor, I., & Michie, S. (2020). Developing quality fidelity and engagement measures for complex health interventions. *British Journal of Health Psychology*, *25*(1), 39-60. http://dx.doi.org/10.1111/bjhp.12394

Poling, A., Carr, J. E. & LeBlanc, L. A. (2002). Operant conditioning. *Encyclopedia of Psychotherapy*, 271-287. https://doi.org/10.1016/B0-12-343010-0/00154-9

World Health Organization. (2018). *International statistical classification of diseases and related health problems* (11th ed.). Retrieved from https://icd.who.int/

Wyatt, G., Sikorskii, A., Rahbar, M. H., Victorson, D., & Adams, L. (2010). Intervention fidelity: Aspects of complementary and alternative medicine research. *Cancer Nursing*, *33*(5), 331-342. http://doi.org/10.1097/NCC.0b013e3181d0b4b7

Yairi, E., & Ambrose, N. (2013). Epidemiology of stuttering: 21st century advances. *Journal of Fluency Disorders, 38*(2), 66-87. https://doi.org/10.1016/j.jfludis.2012.11.002

Yairi, E., & Ambrose, N. G. (2005). *Early Childhood Stuttering*. Pro-Ed.

Yaruss, J. S. (2001). Evaluating treatment outcomes for adults who stutter. *Journal of Communication Disorders, 34*, 163–182. http://doi.org/10.1016/S0021-9924(00)00047-2

Yaruss, J. S. (2010). Assessing quality of life in stuttering treatment outcomes research. *Journal of Fluency Disorders, 35*(3), 190–202. http://doi.org/10.1016/j.jfludis.2010.05.010

Yaruss, J. S., & Quesal, R. W. (2004). Stuttering and the international classification of functioning, disability, and health (ICF): An update. *Journal of Communication Disorders, 37*(1), 35–52. http://doi.org/10.1016/S0021-9924(03)00052-2

Yaruss, J. S., Coleman, C. E., & Quesal, R. W. (2012). Stuttering in school-age children: A comprehensive approach to treatment. *Language, Speech, and Hearing Services in Schools, 43*, 536-548. https://doi.org/10.1044/0161-1461(2012/11-0044)

Yaruss, J. S., & Quesal, R. W. (2016). Overall assessment of the speaker's experience of stuttering (OASES): Documenting multiple outcomes in stuttering treatment. *Journal of Fluency Disorders, 49*, 77–87. https://doi.org/10.1016/j.jfludis.2006.02.002

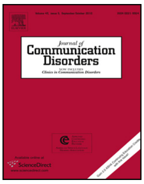
Yeaton, W. H., & Sechrest, L. (1981). Critical dimensions in the choice and maintenance of successful treatments: Strength, integrity, and effectiveness. *Journal of Consulting and Clinical Psychology, 49*(2), 156–167. https://doi.org/10.1037/0022-006X.49.2.156

Paper I

Journal of Communication Disorders 65 (2017) 65–67

Contents lists available at ScienceDirect

Journal of Communication Disorders



Case report

Putting the cart before the horse: A cost effectiveness analysis of treatments for stuttering in young children requires evidence that the treatments analyzed were effective

Íris Ösp Bergþórsdóttir[a], Roger J. Ingham[b,*]

[a] School of Health and School of Education, University of Iceland, Sæmundargötu 2, 101 Reykjavik, Iceland
[b] Department of Speech and Hearing Sciences, University of California, Santa Barbara, Santa Barbara, CA 93106, USA

ARTICLE INFO

Article history:
Received 12 March 2016
Accepted 10 April 2016
Available online 13 April 2016

Keywords:
Stuttering
Treatment
Children

ABSTRACT

Purpose: To investigate the validity of findings from a recent study reported in this journal by de Sonneville-Koedoot, Bouwmans, Franken, and Stolk (2015) on the cost effectiveness of two programs for treating young children who stutter.
Methods: The de Sonneville-Koedoot, Bouwmans et al. study was based directly on the results obtained in an earlier study, known as the RESTART-study, which compared the outcomes from the Lidcombe Program and a Demands and Capacities Model program. The methodology of the RESTART-study was critically reviewed.
Results: The absence of an untreated control group in the RESTART-study makes the results of that study uninterpretable. An inappropriate comparison made with the Yairi and Ambrose (2005) Illinois Study findings failed to resolve the control group problem. Furthermore, the criteria used to classify treated children as "non-stuttering" was also shown to be confounded. The foregoing problems meant that neither treatment program could be shown to be more effective than no treatment.
Conclusion: de Sonneville-Koedoot, Bouwmans et al's findings, which compared the cost effectiveness of two treatments for young children who stutter, have no value for clinical management because the treatments investigated were not shown to be more effective than no treatment.

A recent paper by de Sonneville-Koedoot, Bouwmans, Franken, and Stolk (2015) builds on findings from study by de Sonneville-Koedoot, Stolk, Rietvelt, and Franken (2015) that has come to be known as the RESTART-study. This study was designed to compare the efficacy and/or outcome of the Lidcombe Program (LP) (Onslow, Packman, & Harrison, 2003) and a Program based on the Demands and Capacities model (DCM) (Franken & Putker-de Bruijn, 2014) in the treatment of young children who stutter (CWS). de Sonneville-Koedoot, Stolk et al. (2015, p. 2) claimed that their "results imply that at 18 months post treatment onset, both treatments are roughly equal in treating developmental stuttering in ways that surpass expectations of natural recovery." This is an important and far-reaching conclusion that has already been interpreted by some (e.g., Franken, 2013; Reitzes, 2014) to mean that both treatments are essentially equally effective and can be equally recommended to clinicians. The de Sonneville-Koedoot, Bouwmans et al. (2015) study, which was a cost-management study of the RESTART-study, has however added another twist to this conclusion: that the LP may have an economic advantage over

---

* Corresponding author.
E-mail addresses: iob@hi.is (Í.Ö. Bergþórsdóttir), rjingham@speech.ucsb.edu (R.J. Ingham).

http://dx.doi.org/10.1016/j.jcomdis.2016.04.006
0021-9924/

DCM. As they state, the "cost-effectiveness and cost-utility ratios were in favor of the LP. The LP is considered a good alternative to RESTART-DCM treatment in Dutch primary care." (de Sonneville-Koedoot, Bouwmans et al., 2015, p. 106). But, as we argue below, this conclusion is almost valueless if both programs failed to produce treatment benefits exceeding those that might be achieved without treatment.

Without doubt the most contentious problem with the RESTART-study is that the investigators failed to use an untreated control group to ensure that effects reported for both treatment programs differed from those that might have occurred without treatment. The importance of using untreated control groups in studies investigating the effects of stuttering treatment with young children cannot be overstated. This is because longitudinal studies, especially the Illinois Study (Yairi & Ambrose, 2005), have repeatedly shown very high rates of recovery in the early stages of developmental stuttering. In what follows, therefore, we describe the implications of omitting a control group from the RESTART-study and why it renders their findings uninterpretable. We also address the claim by de Sonneville-Koedoot, Stolk et al. (2015) that the recovery rates in the RESTART study actually exceed those reported by Yairi and Ambrose (2005).

The de Sonneville-Koedoot, Stolk et al. (2015) study is one of the largest ever conducted on 3–6 year old CWS. It involved, ultimately, 198 3–6 y.o CWS with 98 receiving LP (85 completed) and 100 receiving the DC treatment (91 completed). They were assessed at the beginning of treatment and then again at 3, 6, 12 and 18 months after the start of treatment. The investigators concluded that "76.5% (65/85; 95% CI: 66.4–84.2) of children in the LP group were classified as non-stuttering at 18 months compared to 71.4% (65/91; 95% CI: 61.4–79.7) of children in the RESTART-DCM group." (2015, p. 7). Their study is also enriched by an impressive attempt to ensure that the study was conducted with satisfactory levels of fidelity. However the study's result is compromised by a number of unaddressed problems.

The first problem centers on how children were classified as non-stuttering. This was based on an evaluation of "three audio recordings of 10–15 min each in a period of two weeks: one sample of their child speaking to a parent at home, one to a non-family member at home and one to a non-family member away from home." (de Sonneville-Koedoot, Stolk et al., 2015, p. 4). For a child to be classified among "non-stuttering children at 18 months" after the beginning of treatment that child could actually still be producing up to 1.5% syllables stuttered (SS). In other words, by this criterion de Sonneville-Koedoot, Stolk et al. are forced to claim, quite literally, that even though a child in their study was producing up to 1.5% SS at the final evaluation they were not stuttering! The choice of 1.5% SS appears to have been based on false assumptions drawn from two previous studies (Clark, Conture, Walden, & Lambert, 2013; Yaruss, 2000). In one of these (Clark et al., 2013) a low frequency of disfluencies that occur in young normally fluent children were categorized as "stuttered disfluencies", when in fact they were not independently categorized as stuttering events. This mistaken categorization of stuttering among children who were unambiguously classified as "children who do not stutter" leads inexorably to the false assumption that a non-stuttering child may display a low frequency of stuttering. The Yaruss (2000) study is even less pertinent because it did not include a measure of stuttered disfluencies in young children.

That first problem is only magnified by a second when de Sonneville-Koedoot, Stolk et al. (2015) attempt to offset their acknowledged lack of control for "spontaneous recovery" by attempting to compare their findings (see page 14) with data reported by Yairi and Ambrose (2005) on similar aged CWS within the latter's longitudinal Illinois Study. de Sonneville-Koedoot, Stolk et al. (2015) claim, for instance, that in the Illinois Study the recovery rate at 36 months after onset of stuttering was 63% "or higher" which they contend was exceeded by about 10% within their treated CWS at their 18 month evaluation. The comparison is almost akin to comparing apples and oranges. The "recovered" criterion Yairi and Ambrose employed was that "children had to have ceased stuttering for at least 12 months to enter the category of 'recovered"'. Furthermore, Yairi and Ambrose have made it clear that not the slightest sign of stuttering was documented at any point over that 12 month period if a child in the Illinois Study was to be classified recovered or non-stuttering (Yairi & Ambrose, 2005; p. 164; Yairi, personal communication, 10/1/2015). It is clear, therefore, that a child displaying up to 1.5%SS (for an unspecified period) and not being stutter-free for 12 months would never have been classified as recovered or non-stuttering within the Illinois Study. This is underscored by de Sonneville-Koedoot, Stolk et al's claim that almost 74% of their children were categorized "non-stuttering" at 18 months after treatment onset. If the speech performance of those CWS even partially resembled those reported by Yairi and Ambrose – even allowing for a liberal 1.5%SS ceiling for being categorized as non-stuttering – then the RESTART children would need to have been "recovered" since the study's 6 month assessment point. The trend in the mean% SS scores shown in Fig. 2 (2015, p. 12) of de Sonneville-Koedoot, Stolk et al. between the 6 and 18 month assessment points (for either treatment group) is simply incompatible with that possibility—the mean scores for both groups exceeded 2% SS at 6 months and only approached 1% SS by 18 months.

The findings of the RESTART-study, therefore, have actually failed to provide any data that show convincingly that *either* the LP or DCM treatment was effective when compared with no treatment. By contrast there are studies that do show that when LP is used for the treatment of young CWS that it does produce reductions in stuttering that are significantly greater than those obtained by a control group (see Jones et al., 2005; Lattermann, Euler, & Neumann, 2008). However, because of the problems with the design of the RESTART-study it is simply impossible to know if the LP or DCM procedures achieved positive reductions in stuttering in young CWS that exceed those that may have occurred without treatment.

The argument against using an untreated control group in studies on young CWS is that it permits any pernicious effects of untreated stuttering to go unchecked. But there is a relatively simple remedy that de Sonneville-Koedoot, Stolk et al. (2015) could have applied to the design of their study; one that would have largely controlled for the untreated recovery rate problem and avoided an illegitimate comparison with the Yairi and Ambrose (2005) findings. Such a design has also been described by Ingham and Riley (1998). Inspection of Table 1 in the de Sonneville-Koedoot, Stolk et al. (2015) paper shows that approximately 50% of the RESTART children had been stuttering for at least 15 months, a point when Yairi, Ambrose, Paden, and Throneburg (1996) reported that the rate of untreated recovery will dramatically taper off (see also R.Ingham & Cordes, 1999). Consequently, if the speech of these children had been regularly assessed prior to that 15 month point, perhaps at monthly intervals for 3–6 months, then it would have been possible to identify those children whose stuttering frequency was decreasing or not changing. An illustration of such a procedure and the resulting differential performance trends has been provided by Ingham and Riley (1998). Hence the RESTART – study children whose performance was improving might then have been considered reasonably likely to recover without treatment and so able to be distinguished from children who were not improving or in need of treatment – either by LP or by DCM. Comparisons between the baselines of both groups could have then more validly identified treatment-related effects. That relatively simple design change might have considerably reduced the confounding effects of untreated recovery and clarified the comparison between the two treatments. Also, it would have made it possible to interpret the confounded findings from the RESTART-study and, perforce, the necessarily confounded findings of the derivative study reported in this journal by de Sonneville-Koedoot, Bouwmans et al. (2015).

References

Clark, C. E., Conture, E. G., Walden, T. A., & Lambert, W. E. (2013). Speech sound articulation abilities of preschool-age children who stutter. *Journal of Fluency Disorders*, *38*, 325–341.

de Sonneville-Koedoot, C., Bouwmans, C., Franken, M.-C., & Stolk, E. (2015). Economic evaluation of stuttering treatment in preschool children: the RESTART-study. *Journal of Communication Disorders*, *58*, 106–118.

de Sonneville-Koedoot, C., Stolk, E., Rietveld, T., & Franken, M.-C. (2015). Direct versus indirect treatment for preschool children who stutter: the RESTART randomized trial. *PLoS One*, *10*(7), e0133758. http://dx.doi.org/10.1371/journal.pone.0133758.

Franken, M.-C. (2013). Comparing a Demands and Capacities approach and the Lidcombe Program for pre-school stuttering children: the RESTART randomized trial. *Paper read to the national stuttering association conference*.

Franken M-C., & Putker-de Bruijn, D. (2014). RESTART-DCM Method. Treatment protocol developed within the scope of the ZonMW project Cost-effectiveness of the Demands and Capacities Model based treatment compared to the Lidcombe programme of early stuttering intervention: Randomised trial [Internet]. 2014 [cited 2015, January 10]. Available: http://www.nedverstottertherapie.nl/.

Ingham, J. C., & Riley, G. (1998). Guidelines for documentation of treatment efficacy for young children who stutter. *Journal of Speech, Language, and Hearing Research*, *41*, 753–770.

Ingham, R. J., & Cordes, A. K. (1999). On watching a discipline shoot itself in the foot: some observations on current trends in stuttering treatment. In N. B. Ratner, & E. C. Healey (Eds.), *Stuttering research and practice: bridging the gap* (pp. 211–230).Mahwah, NJ: Erlbaum.

Jones, M., Onslow, M., Packman, A., Williams, S., Ormond, T., Schwartz, I., et al. (2005). Randomised controlled trial of the Lidcombe Programme of early stuttering intervention. *British Medical Journal*, *331*, 659–661.

Lattermann, C., Euler, H. A., & Neumann, K. (2008). A randomized control trial to investigate the impact of the Lidcombe Program on early stuttering in German-speaking preschoolers. *Journal of Fluency Disorders*, *33*, 52–65.

Onslow, M., Packman, A., & Harrison, E. (2003). *The Lidcombe Program of early stuttering intervention: a clinician's guide*. Austin, TX: Pro-Ed.

Reitzes, P. (2014). The powered-up parent. *The ASHA Leader*, *19*, 50–56.

Yairi, E., & Ambrose, N. (2005). Early childhood stuttering. *For clinicians by clinicians*. Austin, TX: Pro-Ed Inc..

Yairi, E., Ambrose, N. G., Paden, E. P., & Throneburg, R. N. (1996). Predictive factors of persistence and recovery: pathways of childhood stuttering. *Journal of Communication Disorders*, *29*, 51–77.

Yaruss, J. S. (2000). Converting between word and syllable counts in children's conversational speech samples. *Journal of Fluency Disorders*, *25*, 305–316.

Paper II

CLINICAL LINGUISTICS & PHONETICS
2022, VOL. 36, NO. 10, 904–927
https://doi.org/10.1080/02699206.2021.1965659

Implementation fidelity in parent-implemented interventions for stuttering

Íris Ö. Bergþórsdóttir [a], Kathryn Crowe [a,b], and Jóhanna T. Einarsdóttir [a]

[a]School of Health Science, University of Iceland, Reykjavik, Iceland; [b]School of Teacher Education, Charles Sturt University, Bathurst, Australia

ABSTRACT
Knowledge of the fidelity with which interventions are delivered in research studies is crucial to meaningful examination of intervention impact. This paper presents a review of fidelity implementation (FOI) measurements in interventions jointly delivered by speech-language pathologists (SLP) and parents in research for preschool-aged children who stutter (CWS). Four key FOI components were examined: dosage, adherence, quality, responsiveness. Thirty-six studies met the inclusion criteria for this study. Articles were published between 1990 and 2020 described nine different interventions and examined CWS aged 2–6 years. No study reported all FOI components in both the clinical and the home setting and five did not report on any FOI component in either setting. The number of FOI components reported ranged from 0 to 4 in both clinical (M = 1.5) and home (M = 1.0) settings. Across studies, dosage was most often reported (n = 27, 75.0%) and responsiveness was least often reported (n = 16, 44.4%). The number of FOI components reported in articles did not increase over time, although a trend towards greater reporting in recent years was observed. Poor reporting of FOI in intervention research presents a serious methodological concern that impacts the ability of clinicians and researchers to interpret the findings of these studies. Rigorous measurement and reporting of FOI in future intervention studies is required in order to better inform evidence-based practices for interventions with CWS.

ARTICLE HISTORY
Received 26 November 2020
Revised 3 August 2021
Accepted 3 August 2021

KEYWORDS
Fidelity; parent-implemented; treatment; intervention; stuttering; fluency; children

Since evidence-based practice was introduced as a new paradigm in medicine (Guyatt, 1992), it has become an essential component of quality care in all fields of health, including speech-language pathology (e.g., American Speech-Language-Hearing Association, 2004). Evidence-based practice relies on intervention studies that clearly describe an intervention and meticulously measure and report how it was implemented, in order to understand if it is an effective treatment (Perepletchikova et al., 2007). Key to interpretating treatment effects is knowledge that the intervention described in the research was carried out as it was designed to be (Marks & Tolsma, 1986; Perepletchikova & Kazdin, 2005). Fidelity of implementation (FOI) refers to the degree to which an intervention is implemented as intended or planned (Dane & Schneider, 1998; Dusenbury et al., 2003; Perepletchikova & Kazdin, 2005; Prowse et al., 2015). If an intervention is not implemented as intended, it can

---

CONTACT Íris Ö. Bergþórsdóttir iob@hi.is School of Health Science, University of Iceland, Sæmundargötu 2, 101 Reykjavík, Iceland

This article has been corrected with minor changes. These changes do not impact the academic content of the article.
Supplemental data for this article can be accessed on the publisher's website

be difficult to demonstrate that changes in the dependent variable (e.g., stuttering behaviour) are the result of the independent variable (e.g., the intervention) (Gresham et al., 1993). Therefore, when fidelity is lacking in intervention research, internal and external validity are threatened and this leads to situations in which there is no credible way for inferring causality, i.e., that the intervention was the cause of changes in the target behaviour (Gresham et al., 1993; Moncher & Prinz, 1991). If FOI is not considered, incorrect conclusions could be drawn when making decisions about the benefits of an individual intervention and in comparing different intervention approaches in clinical practice (Bergþórsdóttir & Ingham, 2016). It is for this reason that FOI is now a key consideration in guidelines for intervention studies (e.g., SCRIBE: Single-case reporting guideline in behavioural interventions) and research quality evaluation (e.g., Standards for evidence-based practices in special education (Council for Exceptional Children, 2014; Tate et al., 2016). In summary, failure to demonstrate adequate fidelity in research of clinical interventions is a serious omission that impacts the ability of researchers and clinicians to interpret the effectiveness of the intervention and whether the study and intervention should become part of the evidence-base for practice (Kaderavek & Justice, 2010; McCormack et al., 2017).

Two levels of fidelity are important in intervention research, as well as in clinical practice: *intended fidelity* and *achieved fidelity*. Intended fidelity refers to how the intervention 'should' be implemented in an intervention study or therapeutic setting. Kaderavek and Justice (2010) state that the intended fidelity of an intervention should be the same as the "gold-standard" (p. 370), which is the manner of implementing the intervention that prior research has determined results in the intervention having the maximum effect on the target behaviour. Achieved fidelity, on the other hand, is the 'actual' fidelity with which the intervention was implemented in the intervention study or by the clinician and/or parent working with the child. Achieved fidelity will usually be lower than intended fidelity, due to factors that impact the implementation of the intervention, such as illness that prevents the child from attending all sessions of the intervention (Bellg et al., 2004; Hagermoser Sanetti & Kratochwill, 2014). While differences between intended and achieved fidelity are to be expected, the amount of difference is important, with this sometimes holding the key to understanding why an intervention was unsuccessful, or why its effect varied across participants (Bellg et al., 2004; Hagermoser Sanetti & Kratochwill, 2014). Without measuring and reporting achieved fidelity it is impossible to examine intervention outcomes in an informed manner.

There is currently a lack of systematised ways to report and evaluate FOI (Hagermoser Sanetti & Kratochwill, 2014). Measuring fidelity can be challenging, especially if an intervention is being conducted in clinical settings and/or involves indirect service delivery models. Interventions that are delivered by parents or teachers are examples of contexts where fidelity is challenging to manage (Barton & Fettig, 2013; Lieberman-Betz, 2015; McCormack et al., 2017). Lieberman-Betz's (2015) reviewed studies of intervention research for children's language skills showing how demanding this can be to manage.

Research describing interventions for children who stutter (CWS) have shown that a range of different therapeutic methods can be effective (Andrews et al., 2016; De Sonneville-Koedoot et al., 2015). However, it must be kept in mind that in research of CWS the ability to determine the effect of intervention is complicated by high rates of spontaneous recovery (Yairi et al., 2005). CWS experiencing spontaneous recovery in intervention studies can make non-effective treatments appear effective and can greatly inflate the effect of treatment. Equally, treatments that have poor results in intervention studies could be the result of having either fewer CWS in the study who spontaneously recovered or poorer FOI. Therefore, as there are many factors that result in stuttering diminishing in CWS independent of the intervention, clinicians' confidence in results can be easily undermined (Ingham & Riley, 1998). While many of these factors are not under the control of the researcher, clear and explicit measurement of FOI is important, and will assist in the comparison of outcomes across studies and increase confidence in study findings.

While past research has consistently acknowledged that FOI impacts outcomes in intervention research, FOI measurement itself is rarely the focus of critical review in the field. Therefore, the purpose of this paper is not to review the outcomes of interventions for CWS, but to take a step back and examine measurement and reporting of FOI in intervention studies for CWS. One commonly-used intervention paradigm used with young CWS will be the focus of this investigation – studies which evaluate a parent-implemented intervention.

Parent-implemented intervention for CWS

As parent-child interactions in daily activities have a profound influence on child development, parents are frequently involved in the delivery of intervention for diverse developmental disorders (Kaiser & Roberts, 2013; Wainer & Ingersoll, 2013). Many studies have reported that for children with disabilities, interventions which are partly or completely implemented by children's parents tend to be more effective, such as for children with autism spectrum disorder and intellectual disabilities (Ingersoll & Dvortcsak, 2006; Kaiser & Roberts, 2013). Parent-implemented interventions are built upon a triadic intervention model (Lieberman-Betz, 2015; Salisbury & Cushing, 2013). In this model, clinicians teach parents to implement specific intervention strategies that aim to positively change children's behaviours and/or skills (Lieberman-Betz, 2015). The success of the intervention depends not only on the effectiveness of the intervention itself, but on how well the parent learns and uses the strategies, the frequency with which they use it, and the accuracy with which they use it (Carroll et al., 2007; Ingersoll & Dvortcsak, 2006; Lieberman-Betz, 2015; Roberts & Kaiser, 2011). Measurement of FOI is a crucial element of studies which aim to identify the effectiveness of such interventions because the effectiveness of the intervention delivery as well as the intervention itself must be considered (Bellg et al., 2004; Carroll et al., 2007). As interventions with a parent-implemented element are often delivered in more than one setting, i.e., partly in a clinical/educational setting and partly in a home setting, it is important to document the FOI in all the settings the intervention was delivered in and by all agents who delivered the intervention. When FOI is considered in all the contexts in which an intervention is used, the implementation of the intervention is likely to become more standardized, and therefore more amendable to research and replication (Mowbray et al., 2003).

Parent-implemented interventions for CWS have assumed greater prominence in the literature over recent years, albeit with sometimes controversial findings concerning their efficacy (Bergþórsdóttir & Ingham, 2016; Bothe et al., 2006; De Sonneville-Koedoot et al., 2015). Treatments for CWS are commonly divided into *direct* and *indirect* approaches. Direct approaches, such as the Lidcombe Program (Onslow et al., 2003), provide children with direct instructions and feedback regarding their speech to improve fluency. Indirect approaches, such as Demands and Capacities Model (Franken et al., 2005) typically focus on modifying the demands that proponents argue can lead to stuttering, such as speaking slowly to give the child more time to speak and to reduce demands on speech motor behaviour (Franken et al., 2005). Both direct and indirect approaches to intervention frequently utilize the triadic intervention model, with the SLP delivering intervention in the clinic and parents delivering the intervention between appointments in the child's home (Millard et al., 2009; Unicomb et al., 2017). While the effectiveness and appropriateness of different intervention approaches are hotly debated within the field by researchers, clinicians, and people who stutter alike (Nippold, 2012; Yaruss et al., 2012), FOI is rarely considered (Packman & Onslow, 2012).

Measuring fidelity of implementation

As FOI impacts on intervention outcomes, research studies require well prepared intervention protocols that monitor FOI strictly. However, there currently exist no guidelines or best practices when it comes to measuring FOI. Perepletchikova and Kazdin (2005) make a strong case for the difficulties of making one-size-fits-all measurements of FOI, stating that "because characteristics of the treatments may differ in treatment components and requirements for competent implementation, integrity measures may be developed specifically for each treatment" (p. 375). Many different labels and dimensions have been suggested for measuring treatment implementation fidelity (for an overview see Hagermoser Sanetti & Kratochwill, 2014). However, across the literature a unified comprehension of what the FOI measurements should entail has emerged. This can be distilled into four main components: the amount of the intervention delivered, the accuracy of the delivery of the intervention, the quality of delivery of the intervention, and the participant responsiveness to the intervention (Bellg et al., 2004; Carroll et al., 2007; Dane & Schneider, 1998; Dusenbury et al., 2003; Ingersoll & Dvortcsak, 2006; Lieberman-Betz, 2015; Roberts & Kaiser, 2011). The Appendix presents an overview of the terms and definition for FOI which have been described by specialists in this topic.

In the present paper four components are used, based on a synthesis of previous literature on FOI: dosage, adherence, quality, and responsiveness. These four components encapsulate the definitions of the alternatives presented in the Appendix. Within this paper, these FOI elements are defined as follows. *Dosage* is used to refer to how often the intervention was delivered. This can be shown by the number of sessions attended, session length, and the length of an intervention period in days/weeks/months (Bellg et al., 2004; Dane & Schneider, 1998; Dusenbury et al., 2003; Kaderavek & Justice, 2010; Lieberman-Betz, 2015; Slaughter et al., 2015). This is often measured by keeping a log of sessions attended and/or recordings of sessions. *Adherence* is the accuracy with which the intervention was delivered and is often measured using checklists, automatic recording devices, and/or independent observer judgment (Bellg et al., 2004; Dane & Schneider, 1998; Dusenbury et al., 2003; Kaderavek & Justice, 2010; Lieberman-Betz, 2015; Perepletchikova et al., 2007). *Quality* is how well the intervention is delivered (Bellg et al., 2004; Dane & Schneider, 1998; Dusenbury et al., 2003; Kaderavek & Justice, 2010; Lieberman-Betz, 2015; Perepletchikova et al., 2007) and includes the skill with which the intervention was delivered by the parent/clinician (Gearing et al., 2011), the adults' abilities to engage the child (Santacroce et al., 2004), and awareness of treatment protocols (Perepletchikova et al., 2007). It is often measured by the same method as adherence; the use of checklists, automatic recording devices, and/or independent observer

Table 1. Study Question, Population, Interventions, Comparisons, Outcomes (PICO), and time, type of study and setting definitions.

Study Component	Current Study
Overall Questions	To review evidence of implementation fidelity in published reports of joint parent-clinician implemented interventions for children who stutter (CWS) in both the clinical and home settings.
Population	Children aged 0 to 6 years with diagnosed stuttering.
Interventions	Interventions examined must have implemented in a clinical setting by a SLP and implemented in a home setting with the same/similar intervention and same child in both settings. Interventions will be limited to behavioural interventions that can be utilized by speech-language pathologists and directly address the behavior of stuttering (i.e., not pharmaceutical or electrophysical interventions).
Comparisons	The comparison of interventions within studies will not be limited in this review. Studies will be included that compare the effect of one intervention to no intervention, an alternate intervention/s, and to the same intervention with a participant group with different characteristics. Self-control comparisons will also be considered, as in single-case research designs and crossover designs.
Outcomes	The target outcome in studies will be changes to the behavior of stuttering in children's speech production.
Time	Studies published between 1982 and 2020 will be considered.
Type of Study	Only intervention studies will be considered in this research. These will be studies that provide pre- and post-intervention data for the target outcomes (e.g., experimental group comparisons) or continual data for the target outcomes (e.g., single-case experimental designs).
Setting	Interventions that are delivered in clinical and home settings will be included.

judgment. Finally, *responsiveness* is involvement and engagement in the intervention, based on parent and/or child satisfaction, and/or direct observation of the child's behaviour, as the child is the one the treatment is focused on (Bellg et al., 2004; Century et al., 2010; Dane & Schneider, 1998; Dusenbury et al., 2003; Lieberman-Betz, 2015; Slaughter et al., 2015).

Research aims

To be able to draw conclusions about the effectiveness of an intervention, it is important that the intervention is delivered as prescribed and that deviations from this plan are reported and quantified. Given the lack of prior focus of fidelity in interventions used with CWS, this paper aims to review reporting of four key FOI components in intervention studies of parent-implemented intervention for preschool-aged CWS. The research questions addressed were: (a) How often are the four FOI components reported for the delivery of intervention by clinicians and by parents to CWS? and (b) Has the reporting of FOI components changed over time? A Population, Interventions, Comparisons, Outcomes (PICO) chart of definitions used in this paper is presented in Table 1.

The intention of this paper was not to draw conclusions about the efficacy of interventions used in these studies, but rather to evaluate the extent and quality of FOI measurement and documentation.

Method

Literature search

Systematic search

The search strategy used in this paper was formulated in consultation with a librarian experienced in systematic searching for research. An electronic database search was conducted by the second author through the following databases in February 2021: CINAHL, PubMed, Scopus, and Web of Science. There were four groups of alternate search terms used (OR), with one term from each group required in the search (AND). The terms searched were for population (child OR kid OR pediatric OR pediatrics OR preschool OR pre-school) AND the therapy context (parent OR caregiver OR home) AND the condition (stutter OR stammer OR *fluency) AND intervention (therapy OR intervention OR treatment). Where possible, the search was limited to article abstracts, journal articles, and articles published in English. This search yielded 945 records of which 592 were unique. Search strategies and results for this systematic search are available as Supplemental Material (S1: Search Strategy).

Supplementary searching

Supplemental searching was completed through hand searching and expert advice. Reference lists of included articles were hand searched to identify potentially relevant articles which may have been missed in the systematic database search. The list of included articles was reviewed by three researchers with extensive expertise in the field of stuttering. Nine additional relevant articles were identified in this process.

Selection criteria

The following inclusion criteria were applied hierarchically in determining the eligibility of records for this review. Studies had to (a) be published in English, (b) be published in peer reviewed journals, (c) be published after 1981, (d) describe empirical treatment studies targeting CWS with a behavioural intervention, (e) involve a parent-implemented intervention with CWS, and (f) report on children 6 years of age or younger. The year 1981 was chosen as a limit as this was the year that Yeaton and Sechrest (1981) published an outline of the critical dimensions of treatment fidelity, which arguably marks the beginning of the examination of fidelity in intervention research. As the purpose of this paper was to examine methodology, rather than intervention outcomes, studies utilizing both single-case experimental designs and experimental groups were included.

The first author screened the titles and abstracts of identified articles for eligibility for this review using the inclusion criteria outlined above. 403 records were excluded and 189 not excluded. Two hundred records (33.3%) identified in the systematic search were randomly selected and their titles and abstracts were independently screened by the second author. There was a 100% match between each record's final status (exclude or include) and the screening of the second author (exclude or don't exclude).

Full text screening

Using the inclusion criteria described above, the first author examined the full text of the remaining 189 articles identified in the electronic search and the nine articles identified in the supplemental search. The third author independently screened the full texts of 198 (100%) articles, incorporating the 189 from the systematic search and nine from the supplemental search. Inter-rater reliability was 94%, discrepancy was discussed until consensus was reached. Of the 198 articles screened, 153 articles were excluded, and 36 articles were included. Figure 1 presents the papers selection process.

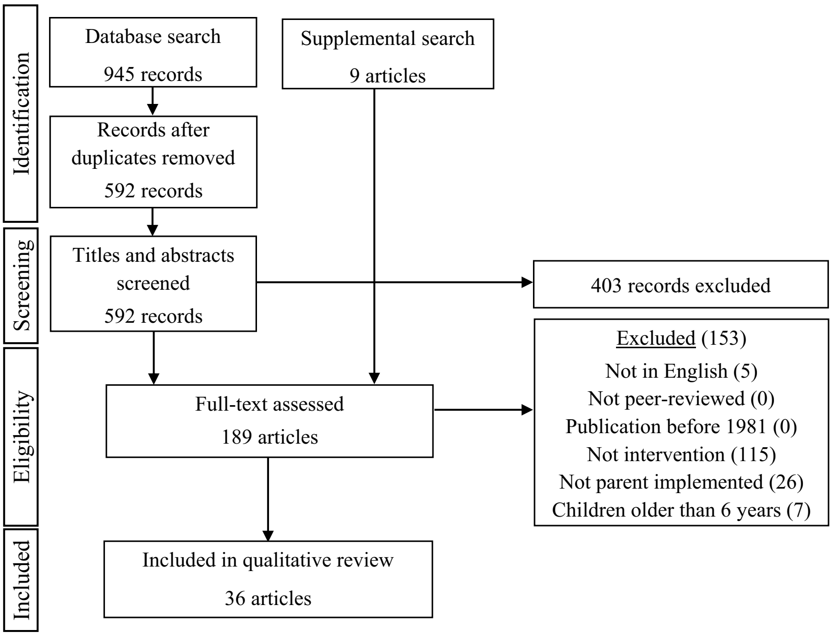


Figure 1. Flowchart of identification, screening, eligibility, and inclusion of articles in this review.

Data extraction and coding

The following information was extracted from each article: publication (year), participant characteristics (sample size, age), methodology (design), and intervention (name/type). Each article was coded as to whether the following FOI components were present/absent for the clinician-implemented intervention and for the parent-implemented intervention: (a) how often the treatment was delivered (*dosage*), (b) the accuracy with which the implementation was delivered (*adherence*), (c) how well the implementation was delivered (*quality*), and (d) the responsiveness of the participants during the intervention (*responsiveness*). The coding scheme applied in this paper was developed by the first and third author, based on review of studies that discussed FOI in the communication science and disorders literature (see Appendix). Each FOI component was defined and coded as follows. *Dosage* was coded as present if at least two of these three criteria were reported: (a) the number of sessions attended, (b) the length of each session or the average length of all sessions, (c) the length of intervention from start to finish. For example, Onslow et al. (1990) refer to the treatment taking place over a specific period (variation between subjects) for the clinical visits (length of clinical sessions recorded and presented) and diary notes of the number of parent sessions, taking place daily. *Adherence* was coded as present if the article referred to any measurement of parent/clinician accuracy in providing the intervention. For example, Druker et al. (2019) describe a logbook to track the implementation of the intervention. *Quality* was coded as present if the article referred on how well the implementation was delivered, e.g., reference to the level of skill, quality of delivery, or engagement. For example, Sawyer et al. (2017) report that video and audio recordings were obtained of the parent/caregiver interacting with the child. *Responsiveness* was coded as present if the article referred to any measurement/report of the child/parent responding to the treatment implementation. For example, Donaghy et al. (2020) report that at both assessments parents were asked about their satisfaction with the improvements.

Coding of FOI variables was binary (present/absent). If an article stated that an aspect of FOI had been evaluated then it was considered to be present, even if the data for the FOI component was not reported in the article. For example, Jones et al. (2005) report that the programme was conducted under the guidance of a speech pathologist and that they gather tape recordings of parents conducting the treatment with their children in everyday speaking environments. However, they report none of the fidelity measurements, only the outcome measurements. The strength of evidence used to report each FOI component was also coded (none, unclear, based on report, evidence/measurement) for the clinician-implemented and the parent-implemented intervention. These codes were defined as follows. *None*: there was no evidence or discussion of this FOI component in the article. *Unclear*: this FOI component was mentioned, but neither the method used to examine this FOI component, nor the outcome of FOI examination was presented in the article. *Based on report*: the articles authors stated that this FOI component was examined, but no data for this FOI component was reported in the article. *Evidence/measurement*: data describing the FOI component in delivery of the intervention was reported in the article. All articles (100.0%) were independently coded for FOI components by the first and third authors. Point-by-point comparison of the data was conducted to determine interrater reliability (following Crowe & Guiberson, 2019), which was 95.3%, with disagreements discussed until consensus was reached.

Results

Study characteristics

A summary of study characteristics for all included articles is presented in Table 2. The 36 studies were published between 1990 and 2020 (*M* = 2011, *SD* = 7.5) and reported 926 CWS. The number of participants in studies ranged from one to 199 children (*M* = 15.5 *SD* = 36.9). Thirty-three studies reported the age range of participants, with participants ranging in age from 2;0 to 6;3 years. Of the remaining three studies, one reported that children were ≤5;11 and two reported the participants' mean ages as 4;2–4;3 and 3.7. Nine different interventions were examined, with some studies investigating variations to standard intervention procedures (e.g., delivery via telehealth). Four studies examined multiple interventions. In order from the most to the least, the interventions examined were the Lidcombe Program (*n* = 25), Parent-Child Interaction Therapy (*n* = 4), Syllable-Timed Interaction (*n* = 3), Operant Program (*n* = 2) and Demands and Capacities Model (*n* = 2). The remaining four interventions were investigated in one article (see Table 2). The research designs used in studies were experimental group comparisons (*n* = 19, 52.8%) or single-case experimental designs (*n* = 17, 47.2%).

Fidelity of implementation

The FOI components reported for implementation of the interventions in clinical settings were as follows: dosage (*n* = 26, 72.2%), adherence (*n* = 7, 19.4%), quality (*n* = 7, 19.4%), and

Table 2. Study characteristics and fidelity of implementation reporting.

Study	N Age Range	Intervention/s (Design)	Dosage C/H	Adherence C/H	Quality C/H	Responsiveness C/H	FOI Score
Onslow et al. (1990)	4 3;2–5;3	Operant Program (SCED)	✓ /✓	-/✓	-/-	-/-	37.5%
Onslow et al. (1994)	23 Mean 3.7	Operant Program (SCED)	✓ /-	-/✓	-/✓	-/✓	50.0%
Matthews et al. (1997)	1 4;2	PCIT (SCED)	-/-	-/-	-/-	-/-	0.0%
Harrison et al. (1999)	1 5;10	Lidcombe Program (SCED)	✓ /-	-/-	✓ /-	-/✓	37.5%
Harris et al. (2002)	23 2;0–4;11	Lidcombe Program (EGC)	-/-	-/-	-/-	-/-	0.0%
Harrison et al. (2004)	38 2;0–5;11	Lidcombe Program (EGC)	-/-	-/-	-/-	-/-	0.0%
Wilson et al. (2004)	5 3;5–5;7	Lidcombe Program (SCED)	✓ /-	-/✓	-/✓	✓ /-	50.0%
Jones et al. (2005)	54 3–6	Lidcombe Program (EGC)	-/-	-/✓	-/✓	-/-	25.0%
Franken et al. (2005)	30 Mean 4;2–4;3	Lidcombe Program Demands and Capacities Model (EGC)	✓ /-	-/-	-/-	✓ /-	25.0%
Lattermann et al. (2005)	4 4;1–5;11	Lidcombe Program (SCED)	-/-	-/-	-/-	-/-	0.0%
Trajkovski et al. (2006)	1 3;2	Syllable-Timed Speech (SCED)	✓ /✓	-/-	-/-	-/-	25.0%
Yaruss et al. (2006)	17 2;6–5;2	FFTP (EGC)	-/-	✓ /-	-/-	✓ /-	25.0%
Lattermann et al. (2008)	46 3;0–5;11	Lidcombe Program (EGC)	✓ /-	-/-	-/-	-/-	12.5%
Lewis et al. (2008)	22 2–6	Lidcombe Program (EGC)	✓ /-	-/✓	-/✓	✓ /-	50.0%
Millard et al. (2008)	6 3;3–4;10	PCIT (SCED)	-/-	-/✓	-/✓	-/-	25.0%
Trajkovski et al. (2009)	3 3;5–3;11	Syllable-Timed Speech (SCED)	✓ /✓	-/-	-/-	-/-	25.0%
Millard et al. (2009)	10 3;7–4;11	PCIT (SCED)	-/-	-/✓	-/✓	-/-	25.0%
Trajkovski et al. (2011)	3 4;7–5;0	Lidcombe Program (SCED)	✓ /-	-/-	-/-	-/-	12.5%
Swift et al. (2011)	1 3–5;9	Syllable-Timed Speech (EGC)	✓ /✓	✓ /✓	✓ /✓	-/✓	87.5%
Femrell et al. (2012)	10 2;9–5;8	Lidcombe Program (SCED)	✓ /-	✓ /-	✓ /-	✓ /-	50.0%
O'Brian et al. (2014)	3 3;6–4;9	Lidcombe Program (SCED)	✓ /-	-/-	-/-	✓ /-	25.0%
Arnott et al. (2014)	54 3;0–5;11	Lidcombe Program (EGC)	✓ /-	✓ /-	-/-	✓ /-	37.5%
De Sonneville-Koedoot et al. (2015)	199 3;0–6;3	Lidcombe Program Demands and Capacities Model (EGC)	✓ /-	✓ /-	✓ /-	-/-	37.5%
Donaghy et al. (2015)	38 2;10–5;10	Lidcombe Program (EGC)	✓ /-	-/-	-/✓	✓ /-	37.5%
Guitar et al. (2015)	14 3;4–6;2	Lidcombe Program (EGC)	✓ /-	-/-	-/-	-/-	12.5%
Bridgman et al. (2016)	19 3;0–5;11	Lidcombe Program (EGC)	✓ /-	-/-	-/-	-/-	12.5%
Swift et al. (2016)	20 2;11–5;2	Lidcombe Program (EGC)	-/✓	-/✓	-/✓	✓ /✓	62.5%
Vong et al. (2016)	4 3;3–4;9	Lidcombe Program (SCED)	✓ /-	-/-	✓ /-	-/✓	37.5%

(*Continued*)

Table 2. (Continued).

Study	N Age Range	Intervention/s (Design)	Dosage C/H	Adherence C/H	Quality C/H	Responsiveness C/H	FOI Score
Sawyer et al. (2017)	17 2;5–5;5	SRSI (EGC)	✓ /✓	✓ /✓	✓ /-	✓ /-	75.0%
Unicomb et al. (2017)	5 3;6–4;7	Lidcombe Program (SCED)	✓ /-	-/-	-/-	-/-	12.5%
Al-Khaledi et al. (2018)	6 3;9–5;11	Lidcombe Program (SCED)	✓ /-	-/✓	-/✓	-/✓	50.0%
Trajkovski et al. (2019)	91 ≤ 5;11	Lidcombe Program Westmead Program (EGC)	-/-	-/-	-/-	-/-	0.0%
Druker et al. (2019)	28 3;0–6;0	Beilby Stuttering Therapy (EGC)	✓ /-	-/✓	-/-	-/-	25.0%
Shafiei et al. (2019)	6 3;0–4;3	Lidcombe Program, PCIT (SCED)	✓ /✓	-/✓	-/✓	-/-	50.0%
Koushik et al. (2019)	31 3;5–5;4	Lidcombe Program (EGC)	✓ /-	-/-	-/-	-/-	12.5%
Donaghy et al. (2020)	89 3;0–5;10	Lidcombe Program (EGC)	✓ /-	✓ /-	✓ /-	✓ /✓	62.5%

C/H = Clinical setting/Home setting; EGC = Experimental group comparison; SCED = Single-case experimental design; INTSS = Intensive smooth speech; EMG = Intensive electromyography feedback; HOMESS = Home-based smooth speech; PCIT = Parent-child interaction therapy; FFTP = Family-focused treatment program; SRSI = Slow relaxed speech intervention.

responsiveness (*n* = 11, 30.6%). The number and percentage of studies reporting FOI components in home sessions were as follows: dosage (*n* = 7, 19.4%), adherence (*n* = 13, 36.1%), quality (*n* = 11, 30.6%), and responsiveness (*n* = 7, 19.4%). The reporting of FOI components in each article is shown in Table 2. The number of FOI components reported ranged from 0 to 4 (*M* = 1.42, SD = 1.18) in clinical settings and 0 to 4 (*M* = 1.03, SD = 1.18) in home settings. For home settings, 17 (47.2%) studies did not include any FOI components, seven (19.4%) studies reported one component, seven (19.4%) studies reported two components, three (8.3%) studies reported three components, and two (5.6%) studies reported all four components. For clinicians, eight (22.2%) studies did not include any FOI components, 14 (38.9%) studies reported one component, eight (22.2%) reported two components, three (8.3%) reported three components, and three (8.3%) reported all four components. Combined reports of FOI in clinical and home settings revealed only 10 (27.8%) studies reporting four or more FOI components and no studies reported all eight. In total, reported FOI components included: dosage (*n* = 33, 45.8%), adherence (*n* = 20, 27.8%), quality (*n* = 18, 25.0%), and responsiveness (*n* = 18, 25.0%). Five of the studies did not report on any FOI components.

The strength of the reporting of each FOI component in the clinical and home settings was evaluated for each article (see Table 3). FOI information on dosage was either presented as evidence or direct measurement (clinic *n* = 26, 72.2%; home *n* = 7, 19.4%) or not mentioned at all. When information on the adherence in studies was presented this was most often in the form of report (clinic *n* = 7, 19.4%; home *n* = 8, 22.2%), but also occasionally as evidence and/or measurement in the home setting (*n* = 2, 5.6%) but not in the clinical setting (*n* = 0, 0.0%). Likewise, quality was also most often based on report (clinic *n* = 5, 13.9%; home *n* = 6, 16.7%) and occasionally as evidence and/or measurement in the home setting (*n* = 3, 8.3%) but not in the clinical setting (*n* = 0, 0.0%). Finally, responsiveness was most often based on report (clinic *n* = 7, 19.4%; home *n* = 4, 11.1%) and

Table 3. Strength of fidelity of implementation reporting across studies (N = 36).

FOI Component	Setting	Not reported	Unclear	Based on report	Evidence/Measurement
Dosage	Clinic	10 (27.8%)	0 (0.0%)	0 (0.0%)	26 (72.2%)
	Home	29 (80.6%)	0 (0.0%)	0 (0.0%)	7 (19.4%)
Adherence	Clinic	29 (80.6%)	0 (0.0%)	7 (19.4%)	0 (0.0%)
	Home	23 (63.9%)	3 (8.3%)	8 (22.2%)	2 (5.6%)
Quality	Clinic	29 (80.6%)	2 (5.6%)	5 (13.9%)	0 (0.0%)
	Home	25 (69.4%)	2 (5.6%)	6 (16.7%)	3 (8.3%)
Responsiveness	Clinic	25 (69.4%)	1 (2.8%)	7 (19.4%)	3 (8.3%)
	Home	29 (80.6%)	3 (8.3%)	4 (11.1%)	0 (0.0%)

occasionally as evidence and/or measurement in the clinical setting (n = 3, 8.3%) but not in the home setting (n = 0, 0.0%).

A Spearman rank-order correlation was conducted to investigate the relationship between year of article publication and the number of FOI components reported. Each of the four components was coded as present/absent for the clinical and the home setting, yielding a possible maximum score of eight. The relationship between publication data and number of FOI components reported was not significant, $r_s(36) = .044$, $p = .80$. Data is visually displayed in Figure 2.

Discussion

The aim of this paper was to review reporting of FOI components in research describing parent-implemented intervention for CWS. In summary, FOI components were more often considered in clinical than in home settings, but overall, reporting of FOI was poor in both settings. Reporting of FOI components rarely involved the reporting of FOI data, with the exception of dosage in clinical settings. Dosage was the FOI component most often reported in clinical settings, followed by responsiveness, adherence and quality. Adherence and quality were the FOI components most often reported in home settings, followed by dosage, and responsiveness. The strength of FOI reporting was the inclusion of evidence of dosage measurement or results of dosage measurement (both for clinician and home settings). Other FOI components were presented as being based on report and in a few cases presented unclearly. However, most FOI components were rarely reported. It was hypothesised that reporting of FOI would increase over time, however, there was no significant change over time, only a weak trend for improvement.

The findings of this study add knowledge of FOI in intervention studies from a new population group to the existing literature describing the generally poor FOI of parent-implemented treatments for children with differences and disabilities (Borelli et al., 2005; Dane & Schneider, 1998; Gresham et al., 1993; Moncher & Prinz, 1991; Perepletchikova et al., 2007; Peterson et al., 1982). The results from Lieberman-Betz's (2015) landmark study on FOI in studies of interventions for children's language skills show both similarities and differences from those of this study. In both studies the FOI component most often reported was dosage for clinicians, and adherence for parents. Similar rates of reporting were found for dosage in clinic (Lieberman-Betz 71%; present study 72.2%) dosage at home (14%; 19.4%) and quality in clinic (20%; 19.4%), and quality at home (29%; 30.6%). However, the present study found lower rates of reporting for the FOI components adherence in clinic (Lieberman-Betz 34%; present study 19.4%), adherence at home (60%; 36.1%),

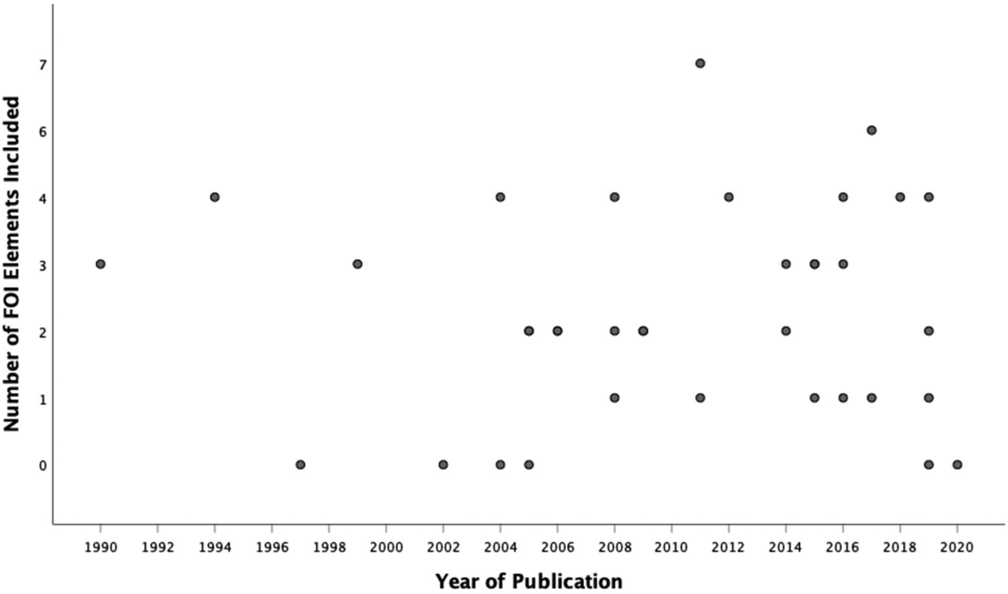


Figure 2. Scatterplot of the relationship between year of study publication and number of FOI components reported in the study.

responsiveness in clinic (54%; 30.6%), and responsiveness at home (43%; 19.4%). The reason for these differences is not clear. Lieberman-Betz (2015) describe a trend within their data for increased reporting of FOI components over time. This trend also occurred in the present study, but it was weak. Although more studies on parent-implemented interventions for CWS were published over time, the average number of FOI components reported did not increase. The difference in reporting of FOI in the two fields requires further investigation to understand why adherence and responsiveness were much less frequently reported in research with CWS.

Authors of all 36 included studies described the research designs, implementation protocols, and outcome measures used in their studies relatively clearly. However, what was absent from most studies was clear information about if and how FOI was monitored and measured. In addition, reporting of FOI data to determine the proximity of intended and achieved fidelity was rarely present. The absence of clear FOI reporting means that clinicians and researchers struggle to examine intervention efficacy and effectiveness systematically and are missing vital information that could contribute to understanding the variance in outcome effects across studies. FOI data can illuminate or eliminate possible reasons for inconsistencies in treatment outcomes between studies (Dane & Schneider, 1998; Dusenbury et al., 2003; Gresham, 1989).

Dosage is an important factor to report, as without attending/administering the intervention there is no treatment. The other components of FOI are equally important and should not be ignored in conducting and reporting treatment studies. Quality and responsiveness could, for example, explain the difference in outcomes where a treatment session was carried out with one-word answers from the child compared to a treatment session where the child engaged in meaningful conversation with multiple opportunities to practice and reinforce the target behaviour. Adherence could explain the difference in outcomes between a treatment session in which the parent provided the required feedback to the child on their behaviour, and a session in which the required feedback is not provided. Therefore, differences in dosage, adherence, quality and/or responsiveness could account for difference in outcomes between children in the same study, between different studies of the same treatment, and between studies of different treatments. Goodhue et al. (2010) noted that mothers in their study said that they "struggle to actually fit it [intervention sessions] in", that they "kept forgetting to do our smooth talking games", and that "it's hard to remember to praise all the time, you forget" (Goodhue et al., 2010, p. 75). Reports such as these are an indication of the importance of conducting systematic FOI measurements to be able to explain the difference in outcome.

Recommendations for future research

Given the findings of this review, the strength or quality of FOI reporting in intervention studies of CWS needs to be improved, in both the clinic and the home setting. Aside from dose in the clinical setting, very few studies provided data reporting on the FOI measured. In place of actual data, authors often made statements acknowledging that one or more aspects of fidelity had been evaluated, or measured, or that the fidelity was satisfactory. In effectiveness research, FOI must be planned, implemented, and reported in order to determine how plausible implementation of the prescribed intervention is in *real-world* situations. This is also a valuable variable to include in analyses examining the differences between children's outcomes in effectiveness trials of the same interventions. When this is done, the consumer of the research is able to see how closely the delivered intervention met the expectations of the planned intervention. FOI measurements in research to inform evidence-based practice requires a stricter framework compared to FOI measurements in conducting practice-based evidence research. However, FOI measurements in both contexts are beneficial.

Researchers should closely follow reporting quality guidelines in planning, implementing, and reporting their research, with particular attention to FOI. A multidimensional guideline is recommended as the best overview of implementation effectiveness (Dane & Schneider, 1998; Lieberman-Betz, 2015; Perepletchikova, 2011; Swift et al., 2011). Such guidelines are available from sources such as the EQUATOR Network (https://www.equator-network.org/) and the Council for Exceptional Children (Council for Exceptional Children, 2014). However, guidelines are not prescriptive in how FOI should be examined in intervention research, leaving researchers with the burden of determining which FOI components to examine. As researchers continually strive to increase the rigor and impact of their work, thorough and clear reporting of multiple FOI components, not only dosage, is essential. Scientific, methodological, and technological advances mean that collection of objective and subjective FOI data in real time within clinical and home settings is a possibility. For example, the use of apps and recording devices in phones can be a feasible choice for researchers to collect FOI data. A good example of FOI reporting can be seen in the article by Swift et al. (2011), they report how many sessions were attended and for how many weeks, they also ask parents to record their beyond-clinic treatments as well as the parents completed a daily diary about the treatment activities. Detailed information on FOI is also becoming a requirement in many journals. In cases where page limits have traditionally inhibited space for providing detailed FOI information within the manuscript, this data can now be included as supplemental online material. Such additions to articles prove beneficial to researchers and clinicians alike.

Clinical implications

The poor reporting of FOI in intervention studies of CWS has important implications for clinicians. Lacking knowledge of the fidelity with which an intervention was delivered in a research study makes use of these studies to inform evidence-based practice questionable (Council for Exceptional Children, 2014). Improvements would give clinicians better insight into the treatment they use in their clinical practice. Measuring FOI could be considered an aspect of best practice for clinicians; however, it also represents a task that clinicians may feel they need to complete on top of their already busy workloads. Clinicians have a responsibility to monitor their client's progress and make changes in intervention if progress is inadequate. Knowledge of how well an intervention is being implemented at home, as well as in the clinic, is crucial to understanding why a client's progress may not be as expected. Measuring FOI can be done by keeping a log of how often the child attends settings and the duration of each setting (most clinicians keep such logs). Within the clinic, this could be done through recording sessions that the clinician then reviews or having another clinician audit the session. At home, measuring FOI is more challenging, with parent diaries, recordings, or logging session in specifically designed apps providing potential easy and implementable solutions. However, Craig et al. (1996) expressed concerns about whether objectively determining the nature of therapeutic input in home environments was possible, as it might be considered intrusive to the point of being unethical. On the other hand, it could be argued that it is also ethically problematic to ask parents to conduct an intervention without even attempting to document that the treatment is being conducted correctly. The clinician needs to balance the demands of parent reporting so as not to induce social acceptability/desirability bias in parent reports of their own intervention and gaining a realistic and data-driven understanding of the actual fidelity with which the intervention is delivered. Accurate fidelity data will lead to more informed clinical decision making.

Limitations

There are limitations to this study that must be borne in mind. Firstly, this review focused only on the FOI data that was published in each study. The justification for this was that only reported data is available to consumers of these research papers. There are plausible reasons that could have led to FOI data being collected but not reported. One reason could be that for the earliest papers included in this review FOI was not a concept that was commonly practiced. Another reason might be journal limits on manuscript length leading to FOI measurements not being included in the manuscript. In addition, journals in the field of speech-language pathology do not demanded demonstrations of FOI in their publishing papers. Secondly, this study only looked at implementation fidelity and did not examine the outcomes of interventions. A key next-step would be to look at more aspects of fidelity, such as fidelity related to analysis of findings (Rietveld, 2020), relationships between FOI reporting, the degree of fidelity attained in the study, and the changes in participants' stuttering behaviour.

Acknowledgments

The authors thank Claudio Dionigi (Charles Sturt University, Australia) for his assistance with the search strategy used in this review. Special thanks to Jack James (Reykjavik University), Roger J. Ingram (University of California, Santa Barbara), and Sigfús Helgi Kristinsson (University of South Carolina) for their assistance with hand searching, helpful comments, critique, and contributions

Disclosure statement

The authors report no conflict of interest.

Funding

This work was supported by the Icelandic Centre for Research grant [141029-052] úbs awarded to J.T. Einarsdóttir

ORCID

Íris Ö. Bergþórsdóttir http://orcid.org/0000-0002-5121-3503
Kathryn Crowe http://orcid.org/0000-0003-3496-5129
Jóhanna T. Einarsdóttir http://orcid.org/0000-0001-8281-5331

References

*Al-Khaledi, M., Lincoln, M., McCabe, P., & Alshatti, T. (2018). The Lidcombe Program: A series of case studies with Kuwaiti preschool children who stutter. *Speech, Language and Hearing*, *21*(4), 224–235. https://doi.org/10.1080/2050571X.2017.1370523

American Speech-Language-Hearing Association. (2004). *Technical report: Evidence-based practice in communication disorders: An introduction*. http://www.asha.org/policy/tr2004-00001/

Andrews, C., O'Brian, S., Onslow, M., Packman, A., & Menzies, R. (2016). Phase II trial of a syllable-timed speech treatment for school-aged children who stutter. *Journal of Fluency Disorders*, *48*, 44–55. https://doi.org/10.1016/j.jfludis.2016.06.001

*Arnott, S., Onslow, M., O'Brian, S., Packman, A., Jones, M., & Block, S. (2014). Group Lidcombe Program treatment of early stuttering: A randomized controlled trial. *Journal of Speech, Language, and Hearing Research*, *57*(5), 1606–1618. https://doi.org/10.1044/2014_JSLHR-S-13-0090

Barton, E. E., & Fettig, A. (2013). Parent-implemented interventions for young children with disabilities: A review of fidelity features. *Journal of Early Intervention*, *35*(2), 194–219. https://doi.org/10.1177/1053815113504625

Bellg, A. J., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D. S., Ory, M., Ogedegbe, G., Orwig, D., & Ernst, D., Czajkowski., & Treatment Fidelity Workgroup of the NIH Behavior Change Consortium. (2004). Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the NIH behavior change Consortium. *Health Psychology*, *23*(5), 443–451. https://doi.org/10.1037/0278-6133.23.5.443

Bergþórsdóttir, I. Ö., & Ingham, R. J. (2016). Putting the cart before the horse: A cost effectiveness analysis of treatments for stuttering in young children requires evidence that the treatments analyzed were effective. *Journal of Communication Disorders*, *65*, 65–67. https://doi.org/10.1016/j.jcomdis.2016.04.006

Borelli, B., Sepinwall, D., Ernst, D., Bellg, A. J., Czajkowski, S., Breger, R., DeFrancesco, C., Levesque, C., Sharp, D. L., Ogedegbe, G., Resnick, B., & Orwig, D. (2005). A new tool to assess treatment fidelity and evaluation of treatment fidelity across 10 years of health behavior research.

Journal of Consulting and Clinical Psychology, *73*(5), 852–860. https://doi.org/10.1037/0022-006X.73.5.852

Bothe, A. K., Davidow, J. H., Bramlett, R. E., & Ingham, R. J. (2006). Stuttering treatment research 1970–2005: Systematic review incorporating trial quality assessment of behavioral, cognitive, and related approaches. *American Journal of Speech-Language Pathology*, *15*(4), 321–341. https://doi.org/10.1044/1058-0360(2006/031)

*Bridgman, K., Onslow, M., O'Brian, S., Jones, M., & Block, S. (2016). Lidcombe Program webcam treatment for early stuttering: A randomized controlled trial. *Journal of Speech, Language, and Hearing Research*, *59*(5), 932–939. https://doi.org/10.1044/2016_JSLHR-S-15-0011

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, *2*(1), 1–9. https://doi.org/10.1186/1748-5908-2-40

Century, J., Rudnick, M., & Freeman, C. (2010). A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Evaluation*, *31*(2), 199–218. https://doi.org/10.1177/1098214010366173

Council for Exceptional Children. (2014). Council for Exceptional Children: Standards for evidence-based practices in special education. *Teaching Exceptional Children*, *46*(6), 206–212. https://doi.org/10.1177/0040059914531389

Craig, A., Hancock, K., Chang, E., McCready, C., Shepley, A., McCaul, A., Costello, D., Harding, S., Kehren, R., Masel, C., & Reilly, K. (1996). A controlled clinical trial for stuttering in persons aged 9 to 14 years. *Journal of Speech, Language, and Hearing Research*, *39*(4), 808–826. https://doi.org/10.1044/jshr.3904.808

Crowe, K., & Guiberson, M. (2019). Evidence-based interventions for learners who are deaf and/or multilingual: A systematic quality review. *American Journal of Speech-Language Pathology*, *28*(3), 964–983. https://doi.org/10.1044/2019_AJSLP-IDLL-19-0003

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, *18*(1), 23–45. https://doi.org/10.1016/S0272-7358(97)00043-3

*Donaghy, M., Harrison, E., O'Brian, S., Menzies, R., Onslow, M., Packman, A., & Jones, M. (2015). An investigation of the role of parental request for self-correction of stuttering in the Lidcombe Program. *International Journal of Speech-Language Pathology*, *17*(5), 511–517. https://doi.org/10.3109/17549507.2015.1016110

*de Sonneville-Koedoot, C., Stolk, E., Rietveld, T., & Franken, M.-C. (2015). Direct versus indirect treatment for preschool children who stutter: The RESTART randomized trial. *PLOS ONE*, *10*(7), e0133758. https://doi.org/10.1371/journal.pone.0133758

*Druker, K. C., Mazzucchelli, T. G., & Beilby, J. M. (2019). An evaluation of an integrated fluency and resilience program for early developmental stuttering disorders. *Journal of Communication Disorders*, *78*, 69–83. https://doi.org/10.1016/j.jcomdis.2019.02.002

*Donaghy, M., O'Brian, S., Onslow, M., Lowe, R., Jones, M., & Menzies, R. G. (2020). Verbal contingencies in the Lidcombe Program: A noninferiority trial. *Journal of Speech, Language, and Hearing Research*, *63*(10), 3419–3431. https://doi.org/10.1044/2020_JSLHR-20-00155 (* Denotes articles included in this review)

Dusenbury, L., Brannigan, R., Falco, F., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implementations for drug abuse prevention in school settings. *Health Education Research*, *18*(2), 237–256. https://doi.org/10.1093/her/18.2.237

*Franken, M. C. J., Kielstra-Van der Schalk, C. J., & Boelens, H. (2005). Experimental treatment of early stuttering: A preliminary study. *Journal of Fluency Disorders*, *30*(3), 189–199. https://doi.org/10.1016/j.jfludis.2005.05.002

*Femrell, L., Åvall, M., & Lindström, E. (2012). Two-year follow-up of the Lidcombe Program in ten Swedish-speaking children. *Folia Phoniatrica Et Logopaedica*, *64*(5), 248–253. https://doi.org/10.1159/000342149

Gearing, R. E., El-Bassel, N., Ghesquiere, A., Badwin, S., Gillies, J., & Ngeow, E. (2011). Major ingredients of fidelity: A review and scientific guide to improving quality of intervention research implementation. *Clinical Psychology Review*, *31*(1), 79–88. https://doi.org/10.1016/j.cpr.2010.09.007

Goodhue, R., Onslow, M., Quine, S., O’Brian, S., & Hearne, A. (2010). The Lidcombe program of early stuttering intervention: Mothers’ experiences. *Journal of Fluency Disorders*, *35*(1), 70–84. https://doi.org/10.1016/j.jfludis.2010.02.002

Gresham, F. M. (1989). Assessment of treatment integrity in school consultation and prereferral intervention. *School Psychology Review*, *18*(1), 37–50. https://doi.org/10.1080/02796015.1989.12085399

Gresham, F. M., Gansle, K. A., Noell, G. H., Cohen, S., & Rosenblum, S. (1993). Treatment integrity of school-based behavioral intervention studies: 1980–1990. *School Psychology Review*, *22*(2), 254–272. https://doi.org/10.1080/02796015.1993.12085651

*Guitar, B., Kazenski, D., Howard, A., Cousins, S. F., Fader, E., & Haskell, P. (2015). Predicting treatment time and long-term outcome of the Lidcombe program: A replication and reanalysis. *American Journal of Speech-Language Pathology*, *24*(3), 533–544. https://doi.org/10.1044/2015_AJSLP-13-0156

Guyatt, G. (1992). Evidence-based medicine: A new approach to teaching the practice of medicine. *Journal of the American Medical Association*, *268*(17), 2420–2425. https://doi.org/10.1001/jama.1992.03490170092032

Hagermoser Sanetti, L. M., & Kratochwill, T. R. (2014). *Treatment integrity: A foundation for evidence-based practice in applied psychology*. American Psychological Association.

*Harris, V., Onslow, M., Packman, A., Harrison, E., & Menzies, R. (2002). An experimental investigation of the impact of the Lidcombe Program on early stuttering. *Journal of Fluency Disorders*, *27*(3), 203–214. https://doi.org/10.1016/S0094-730X(02)00127-4

Harrison, E., Onslow, M., & Menzies, R. (2004). Dismantling the Lidcombe Program of early stuttering intervention: Verbal contingencies for stuttering and clinical measurement. *International Journal of Language and Communication Disorders*, *39*(2), 257–268. https://doi.org/10.1080/13682820310001644551

*Harrison, E., Wilson, L., & Onslow, M. (1999). Distance intervention for early stuttering with the Lidcombe Programme. *Advances in Speech Language Pathology*, *1*(1), 31–36. https://doi.org/10.3109/14417049909167151

Ingersoll, B., & Dvortcsak, A. (2006). Including parent training in the early childhood special education curriculum for children with autism spectrum disorders. *Topics in Early Childhood Special Education*, *26*(3), 179–187. https://doi.org/10.1177/02711214060260030501

Ingham, J. C., & Riley, G. (1998). Guidelines for documentation of treatment efficacy for young children who stutter. *Journal of Speech, Language, and Hearing Research*, *41*(4), 753–770. https://doi.org/10.1044/jslhr.4104.753

*Jones, M., Onslow, M., Packman, A., Williams, S., Ormand, T., Schwarz, I., & Gebski, V. (2005). Randomised controlled trial of the Lidcombe programme of early stuttering intervention. *British Medical Journal*, *331*(7518), 659–661. https://doi.org/10.1136/bmj.38520.451840.E0

Kaderavek, J. N., & Justice, L. M. (2010). Fidelity: An essential component of evidence-based practice in speech-language pathology. *American Journal of Speech-Language Pathology*, *19*(4), 369–379. https://doi.org/10.1044/1058-0360(2010/09-0097)

Kaiser, A. P., & Roberts, M. Y. (2013). Parent-implemented enhanced milieu teaching with preschool children who have intellectual disabilities. *Journal of Speech, Language, and Hearing Research*, *56*(1), 295–309. https://doi.org/10.1044/1092-4388(2012/11-0231)

*Koushik, S., Hewat, S., Onslow, M., Shenker, R., Jones, M., O’Brian, S., Packman, A., Menzies, R., Harrison, E., & Wilson, L. (2019). Three Lidcombe program clinic visit options: A phase II trial. *Journal of Communication Disorders*, *82*, 105919. https://doi.org/10.1016/j.jcomdis.2019.105919

Lattermann, C., Euler, H. A., & Neumann, K. (2008). A randomized control trial to investigate the impact of the Lidcombe Program on early stuttering in German-speaking preschoolers. *Journal of Fluency Disorders*, *33*(1), 52–65. https://doi.org/10.1016/j.jfludis.2007.12.002

*Lattermann, C., Shenker, R. C., & Thordardottir, E. (2005). Progression of language complexity during treatment with the Lidcombe Program for Early Stuttering Intervention. *American Journal of Speech-Language Pathology*, *14*(3), 242–253. https://doi.org/10.1044/1058-0360(2005/024)

*Lewis, C., Packman, A., Onslow, M., Simpson, J., & Jones, M. (2008). A phase II trial of telehealth delivery of The Lidcombe Program of Early Stuttering Intervention. *American Journal of Speech-Language Pathology*, *17*(2), 139–149. https://doi.org/10.1044/1058-0360(2008/014)

Lieberman-Betz, R. G. (2015). A systematic review of fidelity of implementation in parent-mediated early communication intervention. *Topics in Early Childhood Special Education*, *35*(1), 15–27. https://doi.org/10.1177/0271121414557282

Marks, S. E., & Tolsma, R. J. (1986). Empathy research: Some methodological considerations. *Psychotherapy: Theory, Research, Practice, Training*, *23*(1), 4–20. https://doi.org/10.1037/h0085591

*Matthews, S., Williams, R., & Pring, T. (1997). Parent-child interaction therapy and dysfluency: A single-case study. *European Journal of Disorders of Communication*, *32*(3), 346–357. https://doi.org/10.3109/13682829709017900

McCormack, J., Baker, E., Masso, S., Crowe, K., McLeod, S., Wren, Y., & Roulstone, S. E. (2017). Implementation fidelity of a computer-assisted intervention for children with speech sound disorders. *International Journal of Speech-Language Pathology*, *19*(3), 265–276. https://doi.org/10.1080/17549507.2017.1293160

*Millard, S. K., Nicholas, A., & Cook, F. M. (2008). Is parent–child interaction therapy effective in reducing stuttering? *Journal of Speech, Language, and Hearing Research*, *51*(3), 636–650. https://doi.org/10.1044/1092-4388(2008/046)

*Millard, S. K., Edwards, S., & Cook, F. M. (2009). Parent-child interaction therapy: Adding to the evidence. *International Journal of Speech Language Pathology*, *11*(1), 61–76. https://doi.org/10.1080/17549500802603895

Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review*, *11*(3), 247–266. https://doi.org/10.1016/0272-7358(91)90103-2

Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *The American Journal of Evaluation*, *24*(3), 315–340. https://doi.org/10.1177/109821400302400303

Nippold, M. A. (2012). When a school-age child stutters, let's focus on the primary problem. *Language, Speech, and Hearing Services in Schools*, *43*(4), 549–551. https://doi.org/10.1044/0161-1461(2012/12-0054)

O'Brian, S., Smith, K., & Onslow, M. (2014). Webcam delivery of the Lidcombe program for early stuttering: A Phase I clinical trial. *Journal of Speech, Language, and Hearing Research*, *57*(3), 825–830. https://doi.org/10.1044/2014_JSLHR-S-13-0094

*Onslow, M., Costa, L., & Rue, S. (1990). Direct early intervention with stuttering: Some preliminary data. *Journal of Speech and Hearing Disorders*, *55*(3), 405–416. https://doi.org/10.1044/jshd.5503.405

*Onslow, M., Andrews, C., & Lincoln, M. (1994). A control/experimental trial of operant treatment for early stuttering. *Journal of Speech, Language, and Hearing Research*, *37*(6), 1244–1259. https://doi.org/10.1044/jshr.3706.1244

Onslow, M., Packman, A., & Harrison, E. (2003). *The Lidcombe Program of early stuttering intervention: A clinician's guide*. Pro-Ed.

Packman, A., & Onslow, M. (2012). Investigating optimal intervention intensity with the Lidcombe Program of early stuttering intervention. *International Journal of Speech-Language Pathology*, *14*(5), 467–470. https://doi.org/10.3109/17549507.2012.689861

Perepletchikova, F. (2011). On the topic of treatment integrity. *Clinical Psychology: Science and Practice*, *18*(2), 148–153. https://doi.org/10.1111/j.1468-2850.2011.01246.x

Perepletchikova, F., & Kazdin, A. E. (2005). Treatment integrity and therapeutic change: Issues and research recommendations. *Clinical Psychology: Science and Practice*, *12*(4), 365–383. https://doi.org/10.1093/clipsy.bpi045

Perepletchikova, F., Treat, T. A., & Kazdin, A. E. (2007). Treatment integrity in psychotherapy research: Analysis of the studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology*, *75*(6), 829–841. https://doi.org/10.1037/0022-006X.75.6.829

Peterson, L., Homer, A. I., & Wonderlich, S. A. (1982). The integrity of independent variables in behavior analysis. *Journal of Applied Behavior Analysis*, *15*(4), 477–492. https://doi.org/10.1901/jaba.1982.15-477

Prowse, P. D., Nagel, T., Meadows, G. N., & Enticott, J. C. (2015). Treatment fidelity over the last decade in psychosocial clinical trials outcome studies: A systematic review. *Journal of Psychiatry*, *18*(2), 1000258. https://doi.org/10.4172/Psychiatry.1000258

Rietveld, R. (2020). *Human measurement techniques in speech and language pathology: Methods for research and clinical practice*. Routledge.

Roberts, M. Y., & Kaiser, A. P. (2011). The effectiveness of parent-implemented language interventions: A meta-analysis. *American Journal of Speech-Language Pathology*, *20*(3), 180–199. https://doi.org/10.1044/1058-0360(2011/10-0055)

Salisbury, C., & Cushing, L. (2013). Comparison of Triadic and Provider-Led Intervention Practices in Early Intervention Home Visits. *Infants and Young Children*, *26*(1), 28–41. https://doi.org/10.1097/IYC.0b013e3182736fc0

Santacroce, S. J., Maccarelli, L. M., & Grey, M. (2004). Intervention fidelity. *Nursing Research*, *53*(1), 63–66. https://doi.org/10.1097/00006199-200401000-00010

*Sawyer, J., Matteson, C., Ou, H., & Nagase, T. (2017). The effects of parent-focused slow relaxed speech intervention on articulation rate, response time latency, and fluency on preschool children who stutter. *Journal of Speech, Language, and Hearing Research*, *60*(4), 794–809. https://doi.org/10.1044/2016_JSLHR-S-16-0002

*Shafiei, B., Faramarzi, S., Abedi, A., Dehqan, A., & Scherer, R. C. (2019). Effects of the Lidcombe program and parent-child interaction therapy on stuttering reduction in preschool children. *Folia Phoniatrica Et Logopaedica*, *71*(1), 29–41. https://doi.org/10.1159/000493915

Slaughter, S. E., Hill, J. N., & Snelgrove-Clark, E. (2015). What is the extent and quality of documentation and reporting of fidelity to implementation strategies: A scoping review. *Implementation Science*, *10*(1), 129. https://doi.org/10.1186/s13012-015-0320-3

*Swift, M. C., O'Brian, S., Hewat, S., Onslow, M., Packman, A., & Menzies, R. (2011). Investigating parent treatment in the Lidcombe Program. *International Journal of Speech-Language Pathology*, *13*(4), 308–316. https://doi.org/10.3109/17549507.2011.550692

*Swift, M. C., Jones, M., O'Brian, S., Onslow, M., Packman, A., & Menzies, R. (2016). Parent verbal contingencies during the Lidcombe Program: Observations and statistical modeling of the treatment process. *Journal of Fluency Disorders*, *47*, 13–26. https://doi.org/10.1016/j.jfludis.2015.12.002

Tate, R. L., Perdices, M., Rosenkoetter, U., McDonald, S., Togher, L., Shadish, W., Horner, R., Kratochwill, T., Barlow, D. H., Kazdin, A., Sampson, M., Shamseer, L., & Vohra, S. (2016). The Single-Case Reporting Guideline In BEhavioural Interventions (SCRIBE) 2016: Explanation and elaboration. *Archives of Scientific Psychology*, *4*(1), 10–31. https://doi.org/10.1037/arc0000027

*Trajkovski, N., Andrews, C., O'Brian, S., Onslow, M., & Packman, A. (2006). Treating stuttering in a preschool child with syllable timed speech: A case report. *Behaviour Change*, *23*(4), 270–277. https://doi.org/10.1375/bech.23.4.270

*Trajkovski, N., Andrews, C., Onslow, M., Packman, A., O'Brian, S., & Menzies, R. (2009). Using syllable-timed speech to treat preschool children who stutter: A multiple baseline experiment. *Journal of Fluency Disorders*, *34*(1), 1–10. https://doi.org/10.1016/j.jfludis.2009.01.001

*Trajkovski, N., Andrews, C., Onslow, M., O'Brian, S., Packman, A., & Menzies, R. (2011). A phase II trial of the Westmead Program: Syllable-timed speech treatment for pre-school children who stutter. *International Journal of Speech-Language Pathology*, *13*(6), 500–509. https://doi.org/10.3109/17549507.2011.578660

*Vong, E., Wilson, L., & Lincoln, M. (2016). The Lidcombe Program of early stuttering intervention for Malaysian families: Four case studies. *Journal of Fluency Disorders*, *49*, 29–39. https://doi.org/10.1016/j.jfludis.2016.07.003

*Unicomb, R., Hewat, S., Spencer, E., & Harrison, E. (2017). Evidence for the treatment of co-occurring stuttering and speech sound disorder: A clinical case series. *International Journal of Speech-Language Pathology*, *19*(3), 251–264. https://doi.org/10.1080/17549507.2017.1293735

*Trajkovski, N., O'Brian, S., Onslow, M., Packman, A., Lowe, R., Menzies, R., Jones, M., & Reilly, S. (2019). A three-arm randomized controlled trial of Lidcombe Program and Westmead Program early stuttering interventions. *Journal of Fluency Disorders*, *61*, 105708. https://doi.org/10.1016/j.jfludis.2019.105708

Wainer, A., & Ingersoll, B. (2013). Intervention fidelity: An essential component for understanding ASD parent training research and practice. *Clinical Psychology: Science and Practice*, *20*(3), 335–357. https://doi.org/10.1111/cpsp.12045

*Wilson, L., Onslow, M., & Lincoln, M. (2004). Telehealth adaptation of the Lidcombe Program of early stuttering intervention: Five case studies. *American Journal of Speech-Language Pathology*, *13*(1), 81–93. https://doi.org/10.1044/1058-0360(2004/009)

Yairi, E., Ambrose, N., & Grinager, N. (2005). *Early childhood stuttering*. Pro-Ed.

*Yaruss, J. S., Coleman, C., & Hammer, D. (2006). Treating preschool children who stutter: Description and preliminary evaluation of a family-focused treatment approach. *Language, Speech, and Hearing Services in Schools*, *37*(2), 118–136. https://doi.org/10.1044/0161-1461(2006/014)

Yaruss, J. S., Coleman, C. E., & Quesal, R. W. (2012). Stuttering in school-age children: A comprehensive approach to treatment. *Language, Speech, and Hearing Services in Schools*, *43*(4), 536–548. https://doi.org/10.1044/0161-1461(2012/11-0044)

Yeaton, W. H., & Sechrest, L. (1981). Critical dimensions in the choice and maintenance of successful treatments: Strength, integrity, and effectiveness. *Journal of Consulting and Clinical Psychology*, *49*(2), 156–167. https://doi.org/10.1037/0022-006X.49.2.156

Appendix

Components of fidelity of implementation examined in clinical and home intervention sessions and the summary of articles that informed the choice of FOI.

Component	Description	Examples of measurement	Other Terms and Definitions
Dosage: Refers to how much intervention is delivered.	Measurement of how many sessions should be attended/ conducted according to design vs. how many sessions were actually attended/conducted. Measurement of how long each session should be according to design vs. how long each session or average length of session actually was. Measurement of the length of intervention from start to finish	Number of sessions Length of each session Length of intervention	*Dose*: Ensure that treatment "dose" (measured by number, frequency, and length of contact) is adequately described and is the same for each subject within a particular treatment condition. Ensure that treatment dose is the same across conditions, particularly when conditions include multiple behavioral targets (e.g., exercise, smoking) (Bellg et al., 2004). *Dose*: The amount of the program content received by participants. Proportion of intervention providers who received the implementation strategy(s) (i.e., number of people and specific strategy received) (Slaughter et al., 2015). *Dose*: Dose is defined as the amount of program content received by participants. (1) self-reports by providers for all lessons, (2) extrapolations based on observations for a sample of lessons to obtain an objective assessment of the proportion of the curriculum covered and (3) attendance data for each participant (Dusenbury et al., 2003). *Dosage*: Captures the frequency and duration of intervention administration. Active ingredients with regard to dosage may include (a) the number of times the interventionist addresses a target or uses technique during a given treatment session (e.g., 30 models in 30 min), (b) how long a treatment session should last, (c) how often treatment should be delivered throughout a week or month (e.g., 1 hr/week), and (d) the total length of required intervention across time (e.g., 9 months) (Kaderavek & Justice, 2010). *Dosage*: Evaluated based on reporting of the amount of training delivered by providers to parents, and the amount of strategy used by parents with their children (Lieberman-Betz, 2015). *Reporting treatment integrity*: Refers to procedures such as reporting numerical data; reporting overall, component, and session integrity; and reporting the implementation of various integrity procedures (Perepletchikova et al., 2007).

(*Continued*)

(Continued).

Component	Description	Examples of measurement	Other Terms and Definitions
Adherence: Refers to the accuracy with which the intervention is delivered.	Measurement of whether the clinician/parent implemented the treatment intervention according to design (e.g., checklist, second observer, recordings, some other prove of observation). Measurement of whether the clinician/parent implemented the treatment implementation consistently (e.g., checklist, second observer, recordings, some other prove of observation).	Did the clinician/parent give the correct intervention at a stuttering event or/and at a non-stuttering event? Did the clinician/parent consistently give the correct intervention at a stuttering event or/and a non-stuttering event? Did the clinician/parent speak more slowly. Did the clinician/parent give the child more time to speak?	*Delivery*: Monitor and control for subject perceptions of nonspecific treatment effects (e.g., perceived warmth, credibility, etc., of therapist/provider) across intervention and control conditions. Ensure that providers in the same condition are delivering the same intervention. Ensure that the treatments are being delivered in the way in which they were conceived with regard to content and treatment dose. Minimize contamination across treatment/ control conditions, especially when implemented by same provider (Bellg et al., 2004). *Adherence*: The extent to which specified program components were delivered as prescribed in program manuals (Dane & Schneider, 1998). *Adherence*: as the extent to which implementation of particular activities and methods is consistent with the way the program is written (Dusenbury et al., 2003). *Adherence*: Evaluated based on reporting of the use of self- or other- completed checklists to track accuracy of implementation, observational coding of parent training sessions and/or parent-child sessions, and/or measurement of parent implementation as a dependent variable (Lieberman-Betz, 2015). *Adherence*: The extent to which implementation of particular activities and methods is consistent with the way the program is written. Specifying the implementation strategy(s) and evidence of the extent to which this/these implementation strategy(s) took place (Slaughter et al., 2015). *Procedure*: Did the interventionist follow right steps (Kaderavek & Justice, 2010). *Assessing treatment integrity*: Refers to the assessment of treatment integrity via direct, indirect, or hybrid strategies; measurement of therapist treatment adherence as well as competence; and employment of integrity measures with good psychometric properties (Perepletchikova et al., 2007).

(*Continued*)

(Continued).

Component	Description	Examples of measurement	Other Terms and Definitions
Quality: Refers to how well the intervention is delivered.	Measurement of whether the clinician/parent manage to engage with the child during the treatment implementation (e.g., checklist, second observer, recordings, some other prove of observation). Measurement of whether the clinician/parent show awareness to the treatment implementation protocol (e.g., checklist, second observer, recordings, some other prove of observation).	Did the clinician/parent manage to engage in a conversation with the child – enforcing a treatment condition? Are there stuttering events that the clinician/parent is not responding to with a correct intervention? Are there situations where the clinician/parent is not conducting slow relaxed manner of conversional speech?	*Quality*: Evaluated based on the reporting of additional information pertaining to skill of delivery, use of a criterion for mastery, and/or frequency or rate of strategy delivery (Lieberman-Betz, 2015). *Quality*: How well was the intervention delivered (Kaderavek & Justice, 2010). *Enactment*: Ensure that participants actually use the cognitive skills provided in the intervention in appropriate life settings. Ensure that participants actually use the behavioral skills provided in the intervention in appropriate life settings (Bellg et al., 2004). *Quality of delivery*: A measure of qualitative aspects of program delivery that are not directly related to the implementation of prescribed content, such as implementor enthusiasm, leader preparedness, global estimates of session effectiveness, and leader attitudes toward program (Dane & Schneider, 1998). *Quality of delivery*: Ratings of provider effectiveness which assess the extent to which a provider approaches a theoretical ideal in terms of delivering program content (Dusenbury et al., 2003). *Evaluating treatment integrity*: Refers to procedures such as ensuring the accuracy of the representation of the obtained integrity data, training of raters, assessing interrater reliability, and controlling for measure reactivity (Perepletchikova et al., 2007).

(*Continued*)

(Continued).

Component	Description	Examples of measurement	Other Terms and Definitions
Responsiveness: Refers to the involvement and engagement of the participants in the intervention.	Measurement of how the child responds to the treatment implementation and/or measurement of the parents understanding of the treatment protocols (e.g., checklist, second observer, recordings, satisfaction questionnaire or some other prove of observation). Measurement of whether the child is overall happy to participate in the treatment and/or whether the parent is satisfied with the overall process (e.g., checklist, second observer, recordings, satisfaction questionnaire or some other prove of observation). Measurement of whether the child/parent feel that the treatment has improved the child's wellbeing and/or whether the child/parent is overall happy to participate in the treatment (e.g., checklist, second observer, recordings, satisfaction questionnaire or some other prove of observation).	Is the parent's attitude towards the intervention positive? Is the child "happy" to participate in the intervention – does the conversation run smoothly? A measurement of whether the child has improved in their social/school environment.	*Receipt*: Ensure that participants understand the information provided in intervention, especially when participants may be cognitively compromised, have a low level of literacy/education, or not be proficient in English. Make sure that participants are able to use the cognitive skills taught in the intervention (e.g., reframing, problem solving, preparing for high-risk situations, etc.). Make sure that participants are able to use the behavioral skills taught in the intervention (e.g., relaxation techniques, food diaries, cigarette refusal skills, etc.) (Bellg et al., 2004). *Participant Responsiveness*: A measure of participant response to program sessions, which may include indicators such as levels of participation and enthusiasm (Dane & Schneider, 1998). *Participant Responsiveness*: Ratings of the extent to which participants are engaged by and involved in the activities and content of the program (Dusenbury et al., 2003). *Participant Responsiveness*: Participant responsiveness was evaluated based on the reporting or observation of engagement of participants during training or treatment sessions. This may include parent attendance data, report of parent satisfaction for the intervention through questionnaires or interviews, and/or parent-report or direct observation of the child as the parent implemented newly learned strategies (Lieberman-Betz, 2015). *Participant Responsiveness*: The extent to which participants are engaged by and involved in the activities and content of the program. Extent to which intervention providers are involved in the development of the implementation strategy, evaluation of the implementation strategy or their receptivity to the implementation strategy and extent of involvement (Slaughter et al., 2015).

(Based on Bellg et al., 2004; Dane & Schneider, 1998; Dusenbury et al., 2003; Kaderavek & Justice, 2010; Lieberman-Betz, 2015; Perepletchikova & Kazdin, 2005; Slaughter et al., 2015).

Paper III

Journal of Fluency Disorders 83 (2025) 106102

Contents lists available at ScienceDirect

Journal of Fluency Disorders

journal homepage: www.elsevier.com/locate/jfludis

A pilot study of stuttering treatment for older children (STOC)

Jóhanna T. Einarsdóttir [a,*,1], Thorlakur Karlsson [a], Íris Ösp Bergþórsdóttir [a], Kirstín Lára Halldórsdóttir [b], Kathryn Crowe [a,c]

[a] University of Iceland, Iceland
[b] Talþjálfun Mosfellsbæjar (Speech Clinic in Mosfellsbær), Iceland
[c] Charles Sturt University, Australia

ARTICLE INFO

Keywords:
Stuttering
Treatment
School-age children
Fluency
Time-out
Intervention
Parent

ABSTRACT

Purpose: This pilot study explored an operant treatment for school-age children, Stuttering Treatment for Older Children (STOC), in increasing quality of life and the fluency of school-aged children who stutter.
Method: The participants were seven children 9–13 years old. A single-case experimental design using multiple baselines was utilized, with participants randomly assigned to baselines of different lengths. STOC is a home-based treatment program that combined time-out with a specific performance-based criteria. Recordings of children's speech were collected routinely as monthly samples (n = 653) and during treatment sessions (n = 1551). Clinicians immediately reviewed and provided feedback on recordings of sessions. Fidelity was measured in terms of dosage, adherence, quality of the delivery, and responsiveness.
Results: For all participants, the STOC intervention showed promising results. Participant-reported quality of life significantly increased for those participants who completed the STOC program. The frequency of stuttering (percentage of syllables stuttered) decreased by an average of 72 % between the initial baseline measurement and later stages. Additionally, the quantity of speech increased by 18 %, speech sounded more natural, with naturalness scores increased by 50 %, and the severity of stuttering decreased by 56 %. Only four participants finished the intervention. This suggests that the benefits of the intervention include both overall greater well-being and less disfluencies.
Conclusions: This pilot study showed increased speech fluency and quality of life among participants who completed the intervention. However, further investigation is crucial, and we encourage the academic community to engage in this research area.

1. Introduction

Stuttering is described in the International Classification of Functioning, Disability and Health (World Health Organization, 2001) as a disruption in the fluency of speech. However, its impact extends far beyond this, affecting many aspects of life (Yaruss & Quesal, 2006). During the school-age years, research has consistently found associations between stuttering and difficulties in emotional,

---

[*] Correspondence to: University of Iceland, School of Health Sciences and School of Education, Nyi Gardur, Saemundargata 12, Reykjavik 102, Iceland.

E-mail address: jeinars@hi.is (J.T. Einarsdóttir).

[1] https://orcid.org/0000-0001-8281-5331

https://doi.org/10.1016/j.jfludis.2025.106102
Received 2 March 2023; Received in revised form 11 December 2024; Accepted 22 January 2025

behavioral, and cognitive aspects of life (Laiho et al., 2022; Tichenor et al., 2022). For school-age children who stutter, participation in classroom activities can be hampered by negative emotions (Daniels et al., 2012) and social anxiety (Blood & Blood, 2016) that can limit their participation, affect their relationships with teachers, and potentially negatively influence their academic performance (Berchiatti et al., 2020). School-age children who stutter face a higher risk of bullying, which has serious and long-term negative consequences (Blood & Blood, 2016). Many persons who stutter have reported negative experiences from their school years and have reflected on these years as the "most difficult period" of their lives (Baxter et al., 2015, p. 60).

1.1. Previous studies describing school-aged children

There are a diverse range of perspectives on what the focus of treatments for stuttering should be for school-aged children (Bloodstein et al., 2021; Guitar, 2014). Historically, successful management of speech has been one of the primary goals of treatments for stuttering (Euler et al., 2021; Nippold & Packman, 2012).

Recently, a number of systematic reviews have focused on fluency outcome measurement and strategies used to enhance the fluency of school-aged children who stutter (Baxter et al., 2015; Brignell et al., 2021; Johnson et al., 2023; Laiho et al., 2022; Mallick et al., 2021). These reviews underscored the need for further research on treatments for school age children who stutter. The latest systematic review on stuttering treatment for school-aged children (Johnson et al., 2023) emphasized that different treatment approaches can effectively decrease stuttering in school-age children. The 67 studies that met the inclusion criteria for their review were published between 1965 and 2023. Five different treatment approaches were identified (operant methods, speech restructuring, operant methods with speech restructuring, machine-driven intervention, and treatments utilizing cognitive behavior therapy), with studies reporting a wide range of treatment effects. Johnson et al. (2023) concluded that overall, the evidence for treatments that enhanced the fluency of school-aged children who stutter were weak, but that operant methods were an avenue that required more investigation for this group. This conclusion was based both on current evidence and on the success of operant methods with preschool-aged children who stutter.

Recently, emphasis has shifted towards treatments fostering effective communication and teaching children who stutter how to manage the impact of stuttering on their everyday lives (Swift & Langevin, 2024; Yaruss et al., 2012). Three of the most recently published treatment studies of school-aged children who stutter (Euler et al., 2021; Johnson et al., 2024; Kohmäscher et al., 2023) focused on large participant samples and investigated effect of treatment on both quality of life as well and fluency. The quality of life was measured by using the Overall Assessment of the Speakers' Experience of Stuttering (OASES; Yaruss & Quesal, 2006). All three studies showed similar increases in quality of life and significant changes in OASES-S scores as the mean pre-treatment scores decreased with all three treatments (Euler et al., 2021; Johnson et al., 2024; Kohmäscher et al., 2023). In addition to considering participants' wellbeing, these studies also examined changes in fluency associated with the treatment. These studies used three different approaches for modifying the speech: speech restructuring (Euler et al., 2021), stuttering modification (Kohmäscher et al., 2023), and operant programs (Johnson et al., 2024). All reported that the participants stuttered less after treatment but revealed significant individual variations in the responsiveness to these treatments. Euler et al. (2021) found that 16.8 % of the 119 participants showed no discernible treatment benefit. Kohmäscher et al. (2023) noted diverse responses to intervention among the 73 participants, as demonstrated by large standard deviations in the fluency outcome measurements (e.g., percentage of syllables stuttered = 4.5 %) and little long-term benefit reported for out-of-clinic measurements. Johnson et al. (2024) reported that 25 (67.6 %) of the 37 participants showed a partial response to treatment six months after starting, and five (13.5 %) showed no response. These individual variations in treatment responsiveness for fluency outcome are a significant challenge to the internal validity of treatment studies, necessitating that researchers and clinicians consider the relevant participant characteristics that may influence treatment response (Baxter et al., 2015).

1.2. Stuttering treatment for older children (STOC)

This pilot study investigated an operant method as treatment approach for school-age children who stutter: ***Stuttering Treatment for Older Children*** (STOC; see Supplementary Online Material). The program, which is unpublished, was developed by Roger Ingham and Irene Seybold at the University of California, Santa Barbara, around 2011 (Roger Ingham, personal communication). The design of the program is based on earlier research utilizing response-contingent time-out to modify behavior, including speech. When applied to stuttering treatment, response-contingent time-out involves requiring the speaker to pause briefly after stuttering (Bothe et al., 2006; Ingham, 1984; James et al., 1989).

Only a single study was identified that investigated the use of a time-out strategy with school-aged children. Wagaman et al. (1993) examined the effect of using time-out in combination with awareness training and a breathing technique contingent on stuttering occurrences. To participate, the children had to show at least 5 % stuttered words during the first interview. All eight participating children showed less than 3 % stuttered words after treatment. The mean percentage change in stuttering from pre- to post-treatment was 89 %, and this change was maintained for three years. Wagaman et al. (1993) did not report on measurements on changes in quality of life for the participants. As used in STOC and Wagaman et al. (1993), awareness training has been a component of stuttering treatment for decades. It can be traced back to the seminal work of Van Riper (1973), who used stuttering modification techniques like *pull-outs* and *cancellations* to replace the moments of stuttering with controlled fluency. Self-imposed time-out has been shown to be effective in enhancing fluency in studies with adults who stutter (James et al., 1989). Hewat et al. (2006) investigated self-imposed time-out in a study with 22 adults and adolescents. More than half the participants reduced the percentage of syllables they stuttered by more than 50 % at the end of the study. Both studies showed individual variation in responsiveness to the treatment (Hewat et al., 2006; James et al., 1989).

STOC shares characteristics with the Lidcombe program, which is recognized as the best available evidence-based treatment for preschool children (Sjøstrand et al., 2021). Similarly to the Lidcombe program, STOC is a parent-based treatment approach based on response contingent stimulation and is implemented in the home environment (Onslow et al., 2021). STOC differs from the Lidcombe program in that no positive verbal reinforcements are utilized in STOC, and the complexity of children's speech production is not reduced in the early stages of intervention to increase the number of fluent utterances produced. In contrast, in STOC, attention is focused on the child developing an explicit awareness of stuttering events and is supported by developing a time-out response to gain control at the moment stuttering occurs.

STOC is structured so that progress through treatment steps/phases is contingent upon meeting specific performance-based criteria. As increased fluency is achieved, the frequency of intervention sessions decreases. Conversely, when speech is not fluent, the frequency of sessions increases. This approach is similar to the Lidcombe Program, where children can only progress from Stage 1 to Stage 2 once they meet specific criteria or low severity of stuttering (Hewat et al., 2020; Johnson et al., 2024; Lincoln et al., 1996). Specific performance-based criteria have also been used in prolonged speech treatments for adults (see Ingham et al., 2012).

1.3. *The current study*

This pilot study investigates operant treatment, STOC, for school-age children who stutter. The critical elements of the study included analyses of individual participants' performance and long-term follow-up after the treatment period. The fidelity of the intervention was also closely monitored and reported in detail by Bergþórsdóttir, Einarsdóttir, Crowe, & Karlsson, 2025. The present study aimed to establish how well participants responded to STOC and investigate changes in stuttering-related quality of life. The following research questions were addressed:

1. Does participation in STOC result in changes in quality of life?
2. Does participation in STOC decrease the frequency of stuttering of school-aged children who stutter during and after intervention?

2. Method

2.1. *Ethics statement*

Ethical approval was obtained from the Icelandic Bioethical Committee 2014 registration number (S14–059).

2.2. *Participants*

Children in Iceland aged 9–13 years were eligible to participate in this study. The inclusion criteria were (a) that at least 3 % of syllables stuttered during oral reading or conversational speaking tasks when observed in a clinical setting and (b) that the child responded to the time-out procedure prior to commencing treatment. These criteria were selected to ensure that only those children who definitely responded to time-out in a clinical setting were included. The time-out trial involved a series of reading and conversation tasks completed with and without time-out. A child was considered to have responded to time-out if there was a clear decrease in stuttered occasions of at least 50 % in the percent of syllables stuttered in sessions with time-out compared to those without time-out (see Supplementary Online Material). Children were excluded from participation if they had a severe language disorder or developmental delay/disorder. Children with other disorders were eligible to participate to attain a more representative sample. Forty-three potential participants were referred to the study, with 14 not considered further as their parents reported that they no longer stuttered ($n = 11$) or were not interested in participating ($n = 3$). The remaining 29 potential participants were tested in a clinical setting to

Table 1
Participant Background Information.

Participant	Age of onset (years)*	Previous treatment	Comorbid diagnoses	Family history	Age at treatment start (years; months)	Stuttering severity (% SS***)
P1	3	Yes	None	Yes	10;8	2.3
P2	10**	Yes	Anxiety, Language disorder	Yes	13;0	6.7
P3	2	Yes	None	No	13;2	8.0
P4	3	Yes	ADHD, Sound speech disorder	Yes	10;6	15.3
P5	6	No	Anxiety	Yes	9;5	4.0
P6	3	No	ADHD, Literacy difficulties	No	10;4	2.9
P7	3	No	Language disorder	Yes	13;9	13.7

Note.
[*] Age of onset was based on parents' report.
[**] P2 was classified with developmental stuttering as there was a family history of stuttering.
[***] %SS = percentage of syllables stuttered

establish if they met the inclusion criteria. Eighteen children were not eligible to participate as they did not meet the requirements of 3 % syllables stuttered during the oral reading or conversational speaking tasks. The 11 remaining children who produced at least 3 % of syllables stuttered responded to the time-out procedure. Four of these 11 children were excluded at the end of the baseline period as they did not provide the required recordings during the baseline phase. Without baseline data, there would be no way to establish any intervention effect. This resulted in seven children participating in this study.

The seven children who participated were 9–13 years of age (*M* = 11;5, *SD* = 1;7) at the beginning of treatment. All participants were male, of Icelandic origin, and spoke Icelandic. P2 was bilingual and did not live in Iceland at the time of the study, but his parents reported that his dominant language was Icelandic. Table 1 presents further background information about each participant. Parents reported that the onset of the stuttering ranged from 2 to 10 years of age (*M* = 4;5, *SD* = 2;8) and that all participants had continued stuttering since the onset, sometimes with periods of severe stuttering. Participants' stuttering severity was measured based on their first baseline recordings across five different speaking situations and ranged from 2.3 % of syllables stuttered to 15.3 % of syllables stuttered (*M* = 7.5 %, *SD* = 5.2 %). Four participants had previously received treatment for stuttering, with more than one year between being discharged from therapy and commencing participation in this study. The previous treatment approaches varied in both design and dosage. P1 met with the speech-language pathologist at school once a month and practiced modifying his speech by reducing his speech rate. P3 and P5 met with the speech-language pathologist at school once a week and practiced breathing techniques in addition to reducing their speech rate. P4 received Lidcombe therapy during his preschool's years, several years before starting the STOC treatment. The remaining three participants had never received treatment. Four participants (P1, P3, P4, P5) completed the STOC period and the follow-up period, delivering complete data. The remaining three participants withdrew from the study before completing the second treatment phase (see Supplementary Online Material).

2.3. *Experimental design*

A single-case experimental design was employed using a multiple baseline and withdrawal design with follow-up: A_1–B_1–A_2–B_2–A_3 (see Table 2). Participants were randomly assigned to baseline phases (A_1) that were either three months (P2, P3, P7), five months (P1, P5, P6), or six months (P4) in length (see Figs. 1 and 2). All participants then moved to a two-month intervention period (B_1). This was followed by a two-month second baseline period (A2). A second treatment block followed (B_2), which varied in duration depending on the performance of each participant (4–16 months). The follow-up phase (A_3) involved collecting samples for nine months after the end of the intervention, once a month for three consecutive months, and one further sample six months later.

2.4. *Intervention*

The study piloted STOC according to the manual (see Supplementary Online Material). Treatment sessions comprised two 10-minute conversation sessions, one in the morning and the other in the afternoon/evening. A parent-imposed time-out was used in the morning, and a self-imposed time-out was used in the afternoon/evening session. Each time the parent or the child identified a moment of stuttering, the child was to pause from speaking for 2–3 seconds and then continue speaking. Trained judges reviewed each treatment recording and provided feedback via email to the parents on every treatment session within 24 hours of the treatment delivery. The judges were trained by the research team, which consisted of the first and the fourth authors, both of whom are speech-language pathologists specializing in stuttering treatment, measurement, and research. The research team met with the judges at least weekly during the study (more frequently at the beginning), to discuss the treatment process and the feedback of the treatment session. The research team contacted the parents when necessary.

The feedback was on the accuracy of the treatment delivery (the quality and support of the conversation), the amount of stuttering, and how the child proceeded according to the specific performance-based criteria in the STOC manual. As described previously, STOC uses specific criteria to set the spacing and frequency of treatment sessions. Changes in session frequency were dependent on participant performance during sessions. If the participant met the defined criteria for fluent speech during treatment sessions, then the treatment frequency decreased. The criteria for fluency were (a) stutter less than 1 % of syllables, (b) score 1–3 on the naturalness scale, and (c) score 1–2 on the severity scale. These conditions had to be met for three days in a row (i.e., in six consecutive treatment sessions). However, session frequency was increased to the previous step in the treatment schedule if the participant did not meet the defined criteria for fluency. The research team judged if the frequency of sessions was to change according to the specific criteria and sent this information to the parents via email.

Table 2

The OASES Total Scores for Pre- and Post-Treatment for the 4 Participants who Finished the Program by Sections of the Test.

OASES Measures	P1		P3		P4		P5	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Section I	2.7	2.3	3.2	2.7	3.1	1.5	3.2	2.2
Section II	2.4	1.8	2.9	3.1	1.4	2.1	1.8	1.4
Section III	2.1	1.3	2.5	2.3	2.3	1.7	2.0	1.0
Section IV	1.8	1.2	1.4	1.6	1.7	1.2	1.3	1.0
Mean	2.4	1.7	2.8	2.5	2.1	1.7	2.1	1.4

Table 3

Summary of Measurements for each Participant.

Participant	Measure	A_1	B_1	B_1	A_2	B_2	B_2	A_3
	Period	Monthly	Monthly	Treatment	Monthly	Monthly	Treatment	Monthly
	Duration range	3–6 months	2 months	2 months	2 months	4–16 months	4–16 months	9 months
P1	Duration	5 months	2 months	2 months	2 months	15 months	15 months	9 months
	Recordings (n)	25	6	84	10	36	64	20
	%SS	2.0	0.2	0.3	0.4	0.3	0.1	1.0
	SPM	192.6	196.3	163.1	195.3	199.0	180.7	217.7
	Naturalness	4.0	1.0	1.7	1.0	1.0	1.0	2.3
	Severity	4.0	1.0	1.2	1.8	1.0	0.3	2.4
P2	Duration	3 months	2 months	2 months	2 months	2 months	2 months	-
	Recordings (n)	15	10	75	9	10	99	-
	%SS	6.3	3.1	1.9	3.1	1.7	1.1	-
	SPM	157.7	152.8	143.0	158.5	166.6	169.4	-
	Naturalness	6.0	4.0	2.6	3.4	3.0	2.4	-
	Severity	6.0	3.0	2.6	3.2	2.8	1.9	-
P3	Duration	3 months	2 months	2 months	2 months	15 months	15 months	9 months
	Recordings (n)	15	10	102	10	75	99	20
	%SS	8.1	1.1	0.5	2.6	0.8	0.0	0.3
	SPM	155.4	186.0	153.1	171.2	191.7	200.4	189.8
	Naturalness	5.7	2.0	1.8	3.0	1.5	1.2	1.0
	Severity	5.7	2.0	1.6	4.1	1.4	1.0	0.7
P4	Duration	6 months	2 months	2 months	2 months	16 months	16 months	9 months
	Recordings (n)	30	10	113	10	75	237	20
	%SS	17.7	3.8	1.1	0.2	0.1	0.2	0.2
	SPM	82.6	126.9	165.2	196.2	174.1	172.8	180.6
	Naturalness	9.0	5.0	4.2	6.0	3.3	3.7	3.0
	Severity	9.0	4.0	2.6	1.0	0.3	0.6	0.4
P5	Duration	5 months	2 months	2 months	2 months	11 months	11 months	9 months
	Recordings (n)	22	10	107	9	48	58	19
	%SS	2.5	0.4	0.5	0.5	0.4	0.2	1.0
	SPM	154.9	175.2	188.2	171.8	179.7	180.0	184.6
	Naturalness	3.0	1.0	1.4	1.0	1.0	1.0	1.7
	Severity	3.0	1.0	1.5	1.1	0.9	0.7	2.0
P6	Duration	5 months	2 months	2 months	2 months	4 months	5 months	-
	Recordings (n)	25	10	71	10	7	119	-
	%SS	2.3	0.9	1.1	1.0	1.6	0.6	-
	SPM	135.5	134.0	151.1	137.7	143.0	158.2	-
	Naturalness	3.0	1.0	1.7	1.0	1.1	1.1	-
	Severity	4.0	2.0	2.7	1.0	1.7	1.1	-
P7	Duration	3 months	2 months	2 months	2 months	11 months	11 months	-
	Recordings (n)	12	9	105	14	42	218	-
	%SS	17.0	15.7	4.3	12.4	3.8	1.2	-
	SPM	87.2	94.8	125.8	118.0	157.3	164.7	-
	Naturalness	7.0	7.0	4.0	5.4	4.0	2.0	-
	Severity	7.0	7.0	4.5	6.1	4.2	2.2	-
Total	Recordings (n)	144	65	657	72	293	894	79
Mean	%SS	8.0	3.6	1.4	2.9	1.3	0.5	0.6
	SPM	138.0	152.3	155.6	164.1	173.1	175.2	193.2
	Naturalness	5.4	3.0	2.5	3.0	2.1	1.8	2.0
	Severity	5.5	2.9	2.4	2.6	1.8	1.0	1.4

Note. - = no recordings returned. Higher SPM (syllables per minute) means more speech produced. A lower %SS (percentage of syllables stuttered) means fewer syllables stuttered. A lower severity score indicates less severe stuttering. A lower naturalness score means greater speech naturalness.

2.5. Outcome measurements

2.5.1. Changes in quality of life before and after intervention

To assess participants' experiences of stuttering and the impact of this on their quality of life, an Icelandic translation of the OASES-S (Leósdóttir, 2014) was used. OASES-S was designed for use with children aged 7–12 years. The scale has four sections: general information, reactions to stuttering, communication in daily situations, and quality of life (Yaruss, 2010; Yaruss & Quesal, 2006). The participants filled out the Speech Performance Questionnaire and the OASES-S at home before the baseline phases (A_1) started and also after the end of the last phase (A_3).

2.5.2. Speech fluency measures: monthly measures

Monthly video recordings were collected to evaluate the effects of treatment in different speaking situations and to control for variability in speech performance in the same tasks before, during, and after treatment (see Constantino et al., 2016; Ingham et al., 2012; Tichenor & Yaruss, 2021). Speech samples were recorded on iPods and obtained from the participants every month for all five

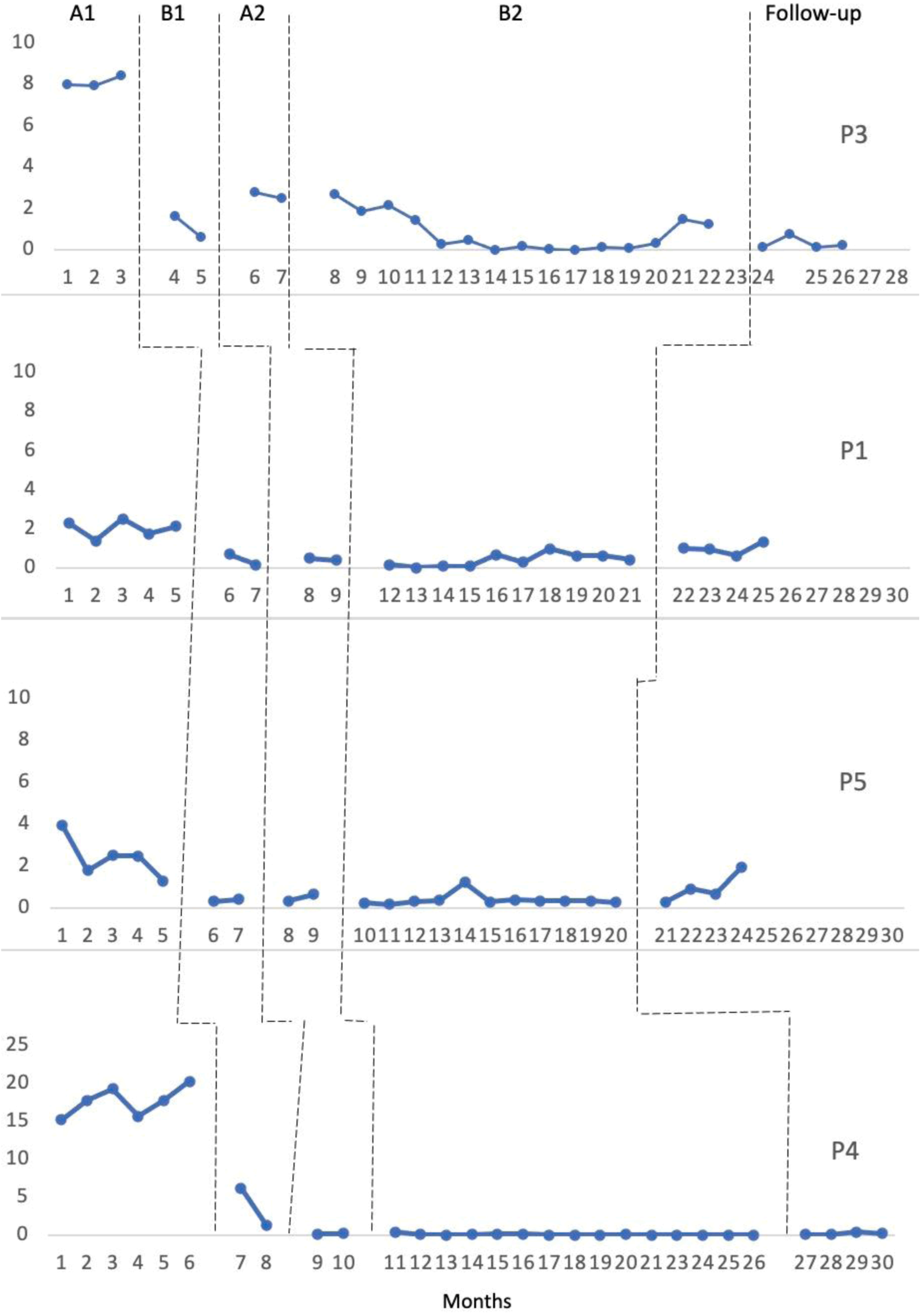


Fig. 1. *Mean percentage of syllables stuttered in Baseline (A) and Treatment (B) for the Four Participants who Finished the Program. Note.* Percentage of syllables stuttered is on the Y-axis, and the range is 0–10 for P3, P1, and P5; however, 0–25 for P4 is an average of the five different speaking situations.

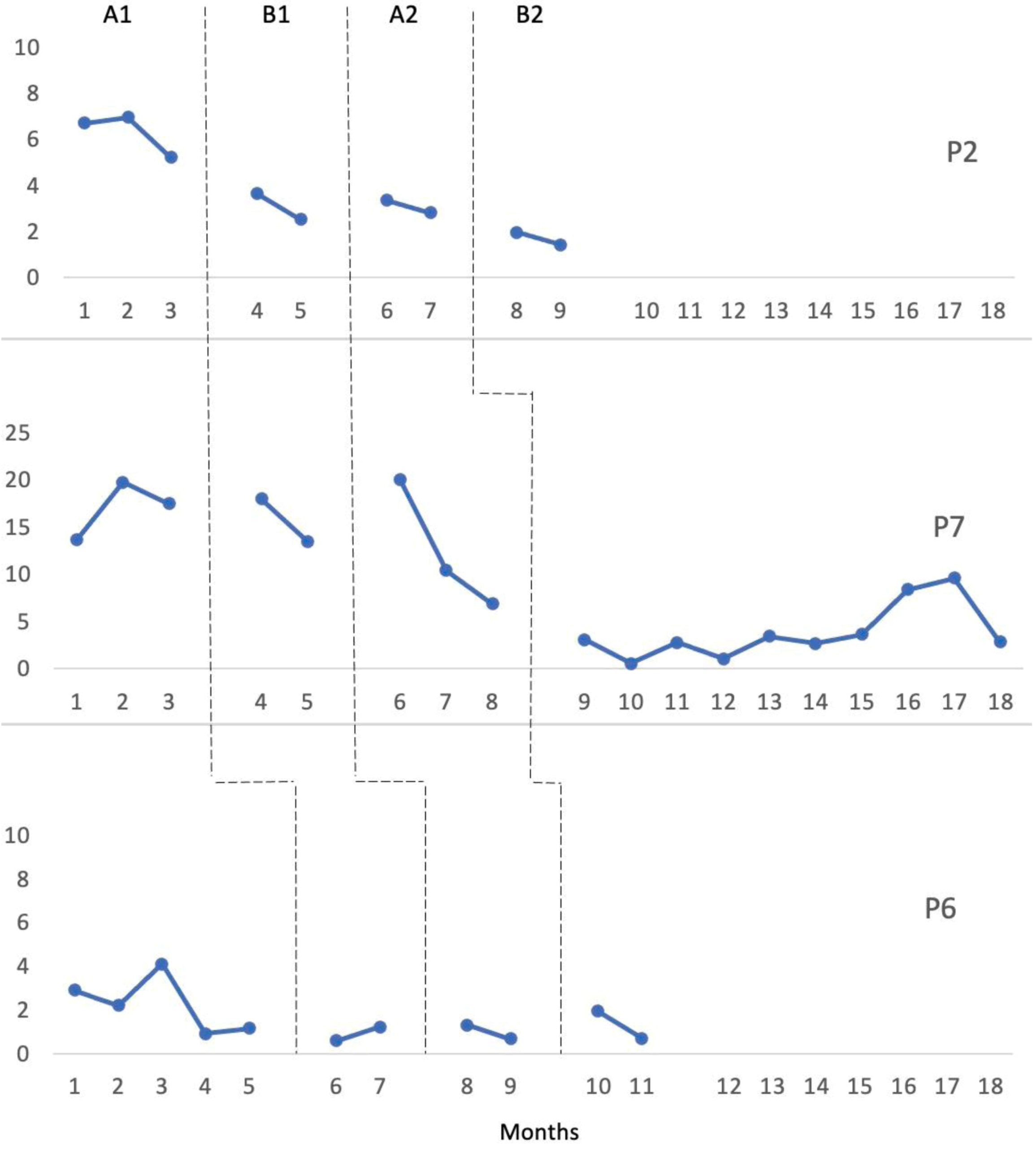


Fig. 2. *Mean percentage of syllables stuttered in Baseline (A) and Treatment (B) for the Three Participants who did not Finish the Program. Note*. Percentage of syllables stuttered is on the Y-axis and the range is 0–10 for P2 and P6, however, 0–25 for P7, and is an average of the five different speaking situations.

study phases (A_1–B_1–A_2–B_2–A_3). Each participant had a shared Google drive with the research team to submit the monthly recordings. Recordings of five different speaking situations were collected and examined monthly from each participant. Recordings were: (a) a three-minute passage read aloud, (b) a three-minute monologue, (c) a ten-minute conversation with friends or relatives, (d) a ten-minute telephone conversation, and (e) a ten-minute, self-selected speaking situation in which the participant determined was challenging for them. Therefore, 36 minutes of recorded speech were collected from each participant every month. Each participant and their parents selected the passage that was read aloud to match each child's reading level and individual interests. Detailed information is available in the Supplementary online material.

The collection of monthly speech samples began in May 2014 and ended with the last follow-up recording in June 2017 – three years and one month in total. Over the course of the study, 653 monthly recordings were obtained from the seven participants, making almost 80 hours of recordings (see Table 2). All recordings were evaluated by following metrics: the number of syllables, the number of stuttering occurrences, syllables per minute, percentage of syllables stuttered, severity scale rating (see Appendix; Einarsdóttir et al., 2014), and naturalness scale rating (Martin et al., 1984).

2.5.3. Speech fluency measures: treatment session measures

Each treatment session was video recorded on an iPod Touch using a custom-designed app. The function of the app was to record the sessions, control for the duration of each treatment session and deliver the recordings directly to the researcher team's iPad. After each session, the parent and the child independently evaluated whether stuttering had occurred during the session and rated its naturalness (Martin et al., 1984). The naturalness scale ranged from 1 (very natural) to 9 very unnatural. The participants judgement on stuttering and their naturalness rating were also sent to the research team through the app. The recording obtained through the app were stored on external drive for further evaluation. The same metrics were evaluated as for the monthly recordings (number of syllables, stuttering occurrences, syllables per minute, percentage of syllables stuttered, severity, and naturalness) using the first three minutes of child speaking time of each treatment session. In total, 1551 treatment recordings (*Range* = 148–350, *M* = 222, *SD* = 81) were received and evaluated for the seven participants (see Table 2).

2.5.4. Changes in participants' perceptions of speech fluency

To evaluate participants' perceptions of stuttering before and after treatment, an Icelandic translation of the Speech Performance Questionnaire (see Boberg & Kully, 1994) was used. The first questionnaire was adapted to this study using five questions from the participants' self-perceptions of fluency and speech. The items used were 1) "Current rating of speech fluency" (terrific, good, fair, poor, terrible), 2) "Now able to sound normal with controlled speech" (never, sometimes, almost always, always), 3) "Now able to speak normally without thinking about controlling speech" (never, sometimes, almost always, always), 4) "Prefer stuttering over controlled speech" (no, sometimes, yes), and 5) "Presently consider myself a stutterer" (yes, no).

2.6. Training, reliability and fidelity measures

2.6.1. Intervention training

Before the first intervention phase (B_1) the participants and their parents were trained to deliver time-out in a clinical setting. The training was conducted by the research team. The training materials were individualized and were based on edited video recordings of each participant's own baseline recordings. The participants and parents were first trained to identify stuttering events on the recordings. Then, during a conversation, the parents practiced stopping the child when a stuttered event occurred, and after that, the child practiced stopping themselves. Training at a conversational level took from 1 to 4 sessions, depending on how long it took for the parents and the child to identify and react to a stuttered event on at least 80 % of stuttering occurrences. If the parents and/or the child repeatedly failed to accurately deliver the treatment during the intervention periods (B_1, B_2), they were retrained in a session delivered either in person or via Skype.

2.6.2. Speech fluency measures: judge training

Seven judges were involved in evaluating participants' stuttering over the course of data collection and data analysis. Judges were the research team and five research assistants who were graduate speech-language pathology students. All judges were trained in assessing stuttering with the Stuttering Measurement System program (see, https://www.hofstra.edu/faculty-staff/faculty-profile.html?id=330, Ingham & Ingham, 2011), the severity scale (see Appendix), and using Icelandic samples on the Stuttering Research and Education Network homepage (https://sturen.vercel.app/ Einarsdóttir et al., 2014). Each judge was trained for 15–20 hours.

2.6.3. Interrater reliability on speech fluency measures

Three judges rescored 88 out of 653 (13.5 %, about every 7th recording) monthly recordings and 116 out of 1551 (7.5 %, about every 13th recording) treatment session recordings to calculate interrater reliability. Systematic random sampling was employed, so in the case of the monthly recordings, a random number generator in Excel was used to find the first recording (a number between 1 and 7), and then every 7th recording was chosen. The same method was employed for treatment session recordings, except, in that case, every 13th recording was chosen. Syllables per minute, percentage of syllables stuttered, naturalness rating, and severity rating were completed for each recording. Monthly recordings were randomly selected across participants and across different speaking situations, with a mean of 13 (*Range* = 7–24) monthly recordings for each participant being selected. Treatment session recordings were also randomly chosen across participants and across morning and afternoon/evening sessions with a mean of 17 treatment recordings (*Range* = 10–23) for each participant selected. Interrater reliability was calculated using the intraclass correlation coefficient for monthly recordings: Syllables per minute (.95), percentage of syllables stuttered (.99), naturalness scores (.91), and stuttering severity scores (.95). For treatment recordings intraclass correlation coefficients were: Syllables per minute (.87), percentage of syllables stuttered (.92), naturalness scores (.78), and stuttering severity scores (.78). Overall, intraclass correlation coefficient was generally high, indicating good reliability (.75–.90) in three cases and excellent reliability (>.90) in five cases (Koo & Li, 2016).

2.6.4. Fidelity measures

Treatment fidelity measures were obtained to determine the treatment delivery. The following four components of fidelity were measured for every treatment session: (a) *dosage* referred to how often the intervention sessions were conducted, (b) *adherence* was evaluated based on how precisely the intervention was delivered, (c) *quality* was evaluated based on how well the intervention was delivered, and (d) *responsiveness* was a judgment of how well the participants engaged in the intervention (Bergþórsdóttir et al., 2022).

For dosage, on average, 80.4 % of the planned treatment sessions were delivered for *Treatment 1* and 86.2 % for *Treatment 2*. The length of the treatment session was controlled by the app and was always 10 minutes, and the segment of the session analyzed was always a three-minute consecutive speech sample of the child's speech from the recording. The average for adherence, or how often the parent/child correctly stopped contingent on stuttering occurrences, in *Treatment 1* for parents was 61.1 % (*Range* = 31.1–78.4) and for children 68.9 % (*Range* = 49.3–90.0). For *Treatment 2*, adherence for parents was 64.3 % (Range = 23.1–92.5), and for children, 70.7 % (*Range* = 18.9–93.9). The wide range, especially for Treatment 2, can be explained by the fact that three participants withdrew from the study during Treatment 2, and these participants had much lower adherence during Treatment 2. The quality (1 = good quality, 4 = poor quality) of the treatment delivery was 1.8 (*Range* = 1.0–2.6) on average in *Treatment 1* and 1.6 (*Range* = 1.1–2.2) in *Treatment 2.* For responsiveness (1 = good conversation, 3 = poor conversation), the average in *Treatment 1* was 1.2 (*Range* = 1.0–1.6) and 1.1 (*Range* = 1.0–1.5) in *Treatment 2*. A detailed analysis of fidelity in this study is presented in Bergþórsdóttir, Einarsdóttir, Crowe, & Karlsson, 2025.

3. Results

3.1. Measures of quality of life

Table 2 shows the changes in OASES-S scores from pre- to post-treatment for the four participants who finished the treatment. Section I includes general information about speech and fluency, Section II reactions and feelings towards stuttering, Section III communication in daily situations, and Section IV questions about quality of life. As can be seen in the table, there was an increase (lower scores) in 13 pre-post cases out of 16. The mean pre-treatment score across sections and participants was 2.24, and the post-treatment score was 1.77 (not shown in Table 3). The difference between the pre- and post-treatment OASES-S scores was significant according to a paired *t*-test, t(15) = 4.1, *p* = .001 (α =.05, one-tailed test).

3.2. Outcome fluency measurements

3.2.1. Outcome fluency measurements – percentage of syllable stuttered

Fig. 1 presents the mean percentage of syllables stuttered in the monthly recordings across all conditions for the four participants who finished the program. All four participants demonstrated less stuttering in monthly recordings during the first treatment condition (B_1) than in the previous baseline (A_1); for example, the mean percentage of syllables stuttered for P1 in the baseline decreased from 2.0 to 0.2 in the treatment condition (90 % drop). A considerable decrease in the percentage of syllables stuttered was observed for participants P3 and P4, who stuttered more severely during the first baseline phase (8.1 and 17.7, respectively, on average) than other participants. However, the mean percentage of syllable stuttered for P3 during the first treatment was 1.1 (86 % drop) and 3.8 for P4 (79 % drop). Also, the mean percentage of syllables stuttered for P5 was 2.5 during the first baseline (A_1), down to 0.4 in the first treatment condition (B_1) – a drop of 84 %. Compared to the first baseline, less stuttering continued for all participants throughout the study, i.e., in B1, A2, B2, and during follow-up (A_3). However, for P3, stuttering increased during the second baseline (mean percentage of syllables stuttered was 2.6 at A_2) and decreased again after four months of treatment (mean percentage of syllables stuttered was 0.8 at B_2).

The results for the participants who did not finish the program are unclear, as data points overlap somewhat between phases – see Fig. 2. In general, however, stuttering decreased from the first baseline (A_1) in the latter conditions for the three who did not finish, showing an effect of STOC. The mean percentage of syllable stuttered for P2 during the first baseline (A_1) was 6.3 and went down to 2.4 on average across the latter conditions (61 % drop); for P6, the mean percentage of syllables stuttered was 2.3 in A_1 and 1.2 in latter conditions (49 % drop), and finally, the mean percentage of syllables stuttered for P7 in A_1 was 17.0 and 7.9 in latter conditions (53 % drop).

The percentage of syllable stuttered (%SS) for all seven participants throughout the study is presented in Table 3, both for the monthly measures and treatment session measures. These are averages for each condition and the follow-up. In all cases, the percentage of syllables stuttered was reduced from the first baseline (A_1) to the latter conditions, as was also observed in Figs. 1 and 2, and this holds up during follow-up for the four participants who were measured. In addition, the treatment measures of percentage of syllables stuttered were lower in the second treatment (B_2) than in the first (B_1) for all participants. Related-Samples Friedman's non-parametric ranks test was employed on the percentage of syllable stuttered data from the monthly recordings. The median in each condition for all seven participants were used to calculate the Friedman's test and it revealed a significant difference in percentage of syllables stuttered between conditions, *Friedman's test statistic* (4) = 10.6, *p* = .016 (α =.05; one-tailed test). The pairwise comparison provided by the same test showed that the significant differences were between the first baseline (A_1) on one hand and each of the latter conditions on the other, i.e., the latter being lower.

3.2.2. Outcome fluency measurements – syllables per minute

In Table 3, syllables per minute is presented for all participants across all conditions for both the monthly and treatment measures. Syllables per minute increased from the first baseline (A_1) to other conditions (all except follow-up [A3]) by 18 % on average across all seven participants. The largest increase was observed for P4 (101 %), then P7 (42 %), P3 (18 %), and P5 (13 %). The increase for the remaining three participants was almost none (1–2 %). The syllables per minute decreased somewhat for participants P2 and P6 between the first baseline and the first treatment condition (B_1). However, syllables per minute increased slightly (3–4 %) on average in conditions A_2 and B_2 above the first baseline level for both participants. A significant difference between conditions was detected

(*Friedman's test statistic* (4) = 11.8, $p = .01$), where pairwise comparison revealed that syllables per minute were significantly higher in B_2 and A_3 (follow-up) than in the first baseline condition (A_1).

3.2.3. Outcome fluency measurements – naturalness and severity

In Table 3, naturalness and stuttering severity scores are presented for all seven participants across all conditions and the monthly and treatment measures. The decrease in scores on average across participants from the first baseline (A_1) to the latter conditions (B_1, A_2, and B_2) was 50 % for naturalness and 56 % for stuttering severity (note that lower scores indicate more naturalness and less stuttering severity). Additionally, the same trend can be observed between the first treatment session (B_1) and the second (B_2), as there was a change towards more naturalness (28 %) and less stuttering severity (56 %) for all participants. A significant difference was detected between conditions for both naturalness (*Friedman's test statistic* (4) = 7.9, $p = .024$) and severity (*Friedman's test statistic* (4) = 9.3, $p = .013$). Pairwise comparison showed significant more naturalness (lower number) in B_1 and B_2 than in A_1, and significantly less severity in A_2 and B_2 than in A_1.

3.2.4. Outcome fluency measurements – speech performance questionnaire

The four participants who finished the program reported changes in speech fluency on the Speech Performance Questionnaire after treatment. For the item "current rating of speech fluency," changes were towards greater speech fluency (pre-treatment: fair $n = 3$, good $n = 1$; post-treatment: good $n = 3$, terrific $n = 1$). When responding to the item "now able to *sound normal* with controlled speech," there was also a shift so that all participants felt they had control (pre-treatment: never $n = 4$; post-treatment: sometimes $n = 1$, almost always $n = 2$, always $n = 1$). When asked about the automaticity of this control in the question "now able to speak normally without thinking about controlling speech," participants reported that less attention was necessary to control their speech (pre-treatment: sometimes $n = 4$; post-treatment: sometimes $n = 1$, almost always $n = 2$, always $n = 1$). Participants were asked about their preferred speech pattern in the question "prefer stuttering over controlled speech" and reported a shift towards preferring controlled speech (pre-treatment: yes $n = 1$, sometimes $n = 1$, no $n = 2$; post-treatment: sometimes $n = 1$, no $n = 3$). Finally, participants were asked whether they currently identified as a person who stuttered with the question "presently consider myself a stutterer," with only one changing their response to indicate that he no longer considered himself a person who stutters (pre-treatment: yes $n = 4$; post-treatment: yes $n = 3$, no $n = 1$).

4. Discussion

This pilot study explored an operant treatment approach for school-age children called STOC mainly by focusing on management of speech for reducing stuttering but also on the participants the quality of life. This parent-based intervention was conducted in a home environment where participants modified their speech during conversation with their parents by using pauses or time-outs contingent on stuttering occurrences during treatment sessions. Parents received intensive guidance and feedback from clinicians, with multiple measurements taken before, during, and after treatment, as well as during long-term follow-up. The intervention positively impacted participant-reported quality of life (OASES-S) for those who completed it and showed increases across fluency measurements or both on increasing fluent speech and decreasing stuttering. The changes in the OASES-S scores from pre- to post-treatment were significant. For all seven participants, the frequency of stuttering (percentage of syllables stuttered) decreased by an average of 72 % between the initial baseline measurement and the latter conditions. Additionally, the quantity of speech (syllables per minute) increased by 18 %; speech sounded more natural, with naturalness scores decreasing by 50 %, and the severity of stuttering decreased by 56 %. The four participants who completed the entire program spoke fluently after the program (1 % syllables stuttered or less), and their speech sounded natural (3 or less on the naturalness rating scale). They reported that they were able to speak without thinking about modifying or controlling their speech most or all the time after the program. The fidelity measurements based on recording of the treatment sessions showed that the parents delivered most of the planned treatment sessions, the correct therapeutic technique was used, the flow of the conversation was generally good, and that the children enjoyed the interaction with their parents during sessions (see Bergþórsdóttir, Einarsdóttir, Crowe, & Karlsson, 2025).

4.1. Participants

The participants in this pilot study were few, with only seven children agreeing to participate and having at least 3 % stuttered syllables. While previous studies have indicated that the effect of time-out on fluency was associated with individual differences (Hewat et al., 2006; Ingham, 1984; James et al., 1989), the inclusion criterion of producing at least 3 percentage of syllables stuttered was selected to ensure that only those children who clearly responded to time-out in a clinical setting were included. This also means that children with less severe stuttering, or 60 % of the potential participants (18 out of 29), were excluded from participating. A previous investigation that employed time-out in connection with breathing techniques for eight school-aged children used a similar inclusion criterion (at least 5 % stuttered words) (Wagaman et al., 1993). Therefore, using time-outs as a strategy has only been investigated in school-aged children who stutter moderately or severely. These studies have limited external validity, as they do not represent the broader population of school-aged children who stutter and most of whom are likely to stutter mildly (Einarsdóttir et al., 2020; 2024; Jones et al., 2006).

In addition, it should be noted that five of the seven participants had been diagnosed with comorbid disorders such as, language disorders, literacy difficulties, ADHD and anxiety. This high rate of comorbidities is in line with previous treatment studies for this age group. For example, in a study by Lincoln et al. (1996), 4 of the 11 participants had other speech and language problems, and in

Andrews et al. (2016), 12 of the 22 participants had a comorbid diagnosis, including speech and language difficulties.

4.2. Quality of life and general improvement

The ultimate goal of stuttering treatment must be a general increase in overall well-being and quality of life. Although the treatment was mainly designed to modify the speech, it also positively affected children's overall communication in daily life. All four sections measured on OASES-S showed mean decreased scores after treatment, comparable to the results reported in the three recent studies on stuttering treatment for school-age children (Euler et al., 2021; Johnson et al., 2024; Kohmäscher et al., 2023). One explanation for the positive change in the quality of life could be that the children experienced greater fluency. Their stuttering decreased and they spoke more both with their parents during the treatment sessions but also in the monthly recordings when speaking in different speaking sessions. The impact of being able to speak more without disfluent speech might have an impact on the participants wellbeing. Another factor could be that the children and the parents talked openly about speaking and stuttering during the intervention. It has been shown that open discussion about stuttering can reduce anxiety (Byrd et al., 2020). The third factor to consider is the extreme commitment and the dedicated time and attention parents give to their children during treatment sessions. The parent's complete focus on the child may have impacted wellness-related outcomes. It is worth noting that certain parents (P2, P7) required guidance on how to engage in supportive conversations with their child to facilitate natural communication. In addition, several parents in this study reported that STOC was associated with positive changes in their child's everyday lives outside of treatment sessions, reporting without prompting that their children's teachers had comments on their child's increased self-esteem and participation in class. Further, fidelity measurements related to quality and responsiveness in the use of STOC showed that the children generally enjoyed the treatment sessions (Bergþórsdóttir, Einarsdóttir, Crowe, & Karlsson, 2025). This indicates that the children did not show frustration or anger while communicating with their parents during the treatment, possible because they experienced decreased stuttered speech.

4.3. Using STOC to modify speaking

STOC was designed as a parent-implemented program utilizing both parent-managed time-out and self-imposed time-out. In this study, this technique was effective in modifying speech and promoting increased fluency, particularly within controlled situations. All seven participants spoke more fluently post-intervention, with the mean percentage change in stuttering from pre-treatment to follow-up being comparable to that reported by Wagaman et al. (1993): 89 % in Wagaman et al. versus 72 % in the current study. As described in the introduction, the three recently published treatment studies of school-aged children who stutter reported less overall change in fluency: in Euler et al. (2021), the decrease was 57 %; in Johnson et al. (2024), it was 67 %, and 25 % in Kohmäscher et al. (2023). It is to note that these three studies all included participants with different severity of stuttering or mild, moderate and sever participants therefore the comparison to STOC could be inflated.

The STOC treatment showed individual variation in the responsiveness to treatment research as documented in many previous stuttering treatments. Although all seven participants clearly responded to the time-out procedure in clinical settings before the study started, three participants (43 %) did not manage to finish the study. A recent investigation of the Lidcombe program with school-aged children who stutter showed a similar outcome, with approximately two-thirds (67 %) of the participants not progressing to Stage 2 in the program (Johnson et al., 2024).

In this pilot study of STOC, all treatment sessions were recorded and evaluated by trained judges. The parents and the children received feedback on the frequency of stuttering and the accuracy of the time-out delivery from researchers after each session. This feedback was particularly crucial during the initial weeks of treatment to ensure the intervention was implemented as intended. Despite this support, some challenges arose in effectively implementing the treatment. These difficulties could be attributed, at least in part, to the neurodiversity of the sample. For instance, P6, who had ADHD, struggled with applying the time-out technique and pausing talking during conversation. Other challenges were related to the lack of interest and support within the family environment, as seen with P7. It is also possible that the children did not like being interrupted during conversation although there were few signs of this on the recordings.

4.4. Time-out versus speech restructuring programs

Previous studies of interventions for school-aged children who stutter that are effective in decreasing the frequency of stuttering have used different variants of speech restructuring, such as prolonged speech, smooth speech, or soft voice onset (Euler et al., 2021) as well as rhythmic speech (Andrews et al., 2016). However, the unnatural speech patterns used in speech restructuring treatments and rhythmic speech could sound strange to children in this age group. Previous research using self-imposed time out for participants older than 14 years of age showed that participants were generally satisfied with the treatment and their speech after treatment (Hewat et al., 2006). The results of this study show that some school-aged children who stutter can achieve fluent speech without the use of speech restructuring, allowing them to sound more natural in their use of fluent speech.

4.5. Limitations

This study has several limitations. The first is regarding external validity. The sample was small, and participants had to meet strict criteria regarding the severity of their stuttering to be eligible for participation. In the inclusion criteria, only children who responded to a time-out condition before the treatment could participate, potentially introducing selection bias. However, all children who were trialled with the time-out condition were responsive to this and, therefore, included in the study. Consequently, we cannot draw definitive conclusions about treatment responsiveness or the generalisability of responsiveness to this treatment for all children who stutter in this age group. Relatedly, there was some self-selection of participants in terms of who completed the study, with the final levels of fluency being unavailable for participants who did not finish. If these data were available, the outcomes of the study could change.

Secondly, the performance requirements for progressing through the program were very demanding, resulting in only four out of the initial seven participants completing the entire treatment program. This indicates that the program, as it is currently, might not be suitable for all children who stutter or their families. Thirdly, the contexts in which participants' fluency was measured in this study lacked real-world settings, so participants' speech was not assessed in natural, real-world contexts. While the monthly recordings were designed to reflect the participants' talking in everyday life, as the participants were so practiced at speaking in these contexts by the end of the study, they may more accurately represent talking in controlled situations. Such a test-retest effect from repeated measurement may have threatened internal validity by inflating estimations of fluency. Finally, there was no documentation of the social validity of STOC, such as could have been attained through interviews with the children, their parents, and teachers about their perspectives on STOC. Data on social validity could perhaps provide more information about why some participants withdrew from STOC while others persisted. These limitations should be considered when interpreting the findings of the STOC pilot study and planning future research designed to evaluate STOC.

4.6. Conclusion

This was a pilot study of operant treatment approach for school-aged children who stutter using time-outs to modify speech. The results were promising in terms of increasing children's speech fluency and in line with previous studies using similar speech modification techniques (Hewat et al., 2006; Wagaman et al., 1993), but they require further investigation. Greater levels of fluency seemed to be associated with psychological measures such as OASES so that more fluent speech could have an impact on the quality of life. Further research with a larger group of participants and additional measures to assess speech in real-world contexts could provide a more comprehensive understanding of STOC's effectiveness and impact.

Funding

The research was funded by the Icelandic Centre for Research (Rannís). The project title: The efficacy of parent-managed response-contingent stuttering treatment program for older aged children who stutter. The Grant number: 141029–052. The PI and the recipient of the grant: Jóhanna Thelma Einarsdóttir.

CRediT authorship contribution statement

Karlsson Thorlakur: Formal analysis, Data curation. **Einarsdottir Johanna Thelma:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition. **Halldórsdóttir Kirstín Lára:** Supervision, Project administration, Investigation. **Bergþórsdóttir Íris:** Writing – original draft, Investigation, Data curation. **Crowe Kathryn:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Data curation.

Declaration of Competing Interest

The authors confirm that there is no conflict of interest concerning the MS or the treatment method described in the MS

Acknowledgments

Special thanks to Roger J. Ingham for his contribution to designing the study and to Jack James for his comments. We would like to thank Michael Blomgren for his helpful comments and for editing the final version of the manuscript. Thank you to the following clinicians and research assistants for their work on the study: Ása Birna Einarsdóttir, Berglind Bjarnadóttir, Bjarnfríður Leósdóttir, Rósa Hauksdóttir, Sigfús Helgi Kristinsson, and Ragnhildur Gunnarsdóttir. We would also like to thank the children and parents who participated in this study. Thank you to the reviewers for their comments, which strengthened our paper.

Appendix. Severity Rating Training

The severity rating scale employed is an 11-point ordinal scale ranging from 0 to 10 to assess the overall severity of stuttering. This scale was established based on agreement among experts within the Stuttering Research and Education Network, which involves researchers and clinicians from Iceland, Denmark, Sweden, Norway, Finland, and Belgium.

This scale provides an overall indication of severity: 0 indicates no stuttering, 1 indicates very mild stuttering, 2 indicates mild stuttering, 5 indicates moderate stuttering, and 10 indicates extremely severe stuttering.

In this training session you will learn to rate stuttering severity based on 33 video samples. The rating of the samples is based on an agreement of six experienced Icelandic speech-language pathologists. They independently rated each sample, and then they discussed the severity rating until they reached a consensus. The samples are organized from 0 to 10. You can click on each severity rating and play the video demonstrating a speech sample rated with stuttering severity. You have the option to view more samples with the same rating.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jfludis.2025.106102.

Data availability

Data will be made available on request.

References

Andrews, C., O'Brian, S., Onslow, M., Packman, A., Menzies, R., & Lowe, R. (2016). Phase II trial of a syllable-timed speech treatment for school-age children who stutter. *Journal of Fluency Disorders, 48*, 44–55. https://doi.org/10.1016/j.jfludis.2016.06.001

Baxter, S., Johnson, M., Blank, L., Cantrell, A., Brumfitt, S., Enderby, P., & Goyder, E. (2015). The state of the art in non-pharmacological interventions for developmental stuttering. Part 1: A systematic review of effectiveness. *International Journal of Language Communication Disorders, 50*(5), 676–718. https://doi.org/10.1111/1460-6984.12171

Berchiatti, M., Badenes-Ribera, L., Ferrer, A., Longobardi, C., & Gastaldi, F. G. M. (2020). School adjustment in children who stutter: The quality of the student-teacher relationship, peer relationships, and children's academic and behavioral competence. *Children and Youth Services Review, 116*, Article 105226. https://doi.org/10.1016/j.childyouth.2020.105226

Bergþórsdóttir, Í .Ö., Crowe, K., & Einarsdóttir, J. T. (2022). Implementation fidelity in parent-implemented interventions for stuttering. *Clinical Linguistics Phonetics, 36*(10), 904–927. https://doi.org/10.1080/02699206.2021.1965659

Blood, G. W., & Blood, I. M. (2016). Long-term consequences of childhood bullying in adults who stutter: Social anxiety, fear of negative evaluation, self-esteem, and satisfaction with life. *Journal of Fluency Disorders, 50*, 72–84. https://doi.org/10.1016/j.jfludis.2016.10.002

Bloodstein, O., Ratner, N. B., & Brundage, S. B. (2021). A handbook on stuttering. *Plural*.

Boberg, E., & Kully, D. (1994). Long-term results of an intensive treatment program for adults and adolescents who stutter. *Journal of Speech and Hearing Research, 37*(5), 1050–1059. https://doi.org/10.1044/jshr.3705.1050

Bothe, A. K., Davido, J. H., Bramlett, R. E., & Ingham, R. J. (2006). Stuttering treatment research. 1970–2005: I. Systematic review incorporating trial quality assessment of behavioral, cognitive, and related approaches. *American Journal of Speech-Language Pathology, 15*(4), 321–341. https://doi.org/10.1044/1058-0360(2006/031)

Brignell, A., Krahe, M., Downes, M., Kefalianos, E., Reilly, S., & Morgan, A. (2021). Interventions for children and adolescents who stutter: A systematic review, meta-analysis, and evidence map. *Journal of Fluency Disorders, 70*, Article 105843. https://doi.org/10.1016/j.jfludis.2021.105843

Byrd, C. T., Werle, D., & St. Louis, K. (2020). Speech-language pathologists' comfort level with the use of the term 'stuttering' during evaluations. *American Journal of Speech-Language Pathology, 29*, 841–850. https://doi.org/10.1044/2020_AJSLP-19-00081

Constantino, C. D., Leslie, P., Quesal, R. W., & Yaruss, J. S. (2016). A preliminary investigation of daily variability of stuttering in adults. *Journal of Communication Disorders, 60*, 39–50. https://doi.org/10.1016/j.jcomdis.2016.02.001

Daniels, D. E., Gabel, R. M., & Hughes, S. (2012). Recounting the K-12 school experiences of adults who stutter: A qualitative analysis. *Journal of Fluency Disorders, 37*(2), 71–82. https://doi.org/10.1016/j.jfludis.2011.12.001

Einarsdóttir, J. T., Crowe, K., Kristinsson, S. H., & Másdóttir, T. (2020). The recovery rate of early stuttering. *Journal of Fluency Disorders, 64*, Article 105764. https://doi.org/10.1016/j.jfludis.2020.105764

Bergþórsdóttir, Í .Ö., Einarsdóttir, J.T., Crowe, K., & Karlsson, T. (2025). The effect of fidelity in STOC stuttering treatment for school-aged children. *Manuscript submitted for publication*.

Einarsdóttir, J.T., Halldórsdóttir K.L., Leósdóttir, B., & Bjarnadóttir, B. (2014). *STUREN homepage-Icelandic samples*. Retrieved from ⟨https://sturen.vercel.app/⟩.

Einarsdóttir, J. T., Hermannsdóttir, B., & Crowe, K. (2024). A prospective 14-year follow-up study of the persistence and recovery of stuttering. *Journal of Fluency Disorders, 80*, Article 106058. https://doi.org/10.1016/j.jfludis.2024.106058

Euler, H. A., Merkel, A., Hente, K., Neef, N., von Gudenberg, A. W., & Neumann, K. (2021). Speech restructuring group treatment for 6-to-9-year-old children who stutter: A therapeutic trial. *Journal of Communication Disorders, 89*, Article 106073. https://doi.org/10.1016/j.jcomdis.2020.106073

Guitar, B. (2014). *Stuttering: An integrated approach to its nature and treatment* (4th ed.). Lippincott Williams & Wilkins.

Hewat, S., Onslow, M., Packman, A., & O'Brian, S. (2006). A phase II clinical trial of self-imposed time-out treatment for stuttering in adults and adolescents. *Disabilities and Rehabilitation, 28*(1), 33–42. https://doi.org/10.1080/09638280500165245

Hewat, S., Unicomb, R., Dean, I., & Cui, G. (2020). Treatment of Childhood stuttering using the Lidcombe Program in mainland China: Case studies. *Speech, Language and Hearing, 23*(2), 55–65. https://doi.org/10.1080/2050571X.2018.1511106

Ingham, R. J. (1984). *Stuttering and behavior therapy: Current status and experimental foundations*. College-Hill Press.

Ingham, J.C. & Ingham, R.J. (2011) The Stuttering Measurement System (SMS) training manual. Available at ⟨https://www.hofstra.edu/faculty-staff/faculty-profile.html?id= 330⟩.

Ingham, R. J., Ingham, J. C., & Bothe, A. K. (2012). Integrating functional measures with treatment: A tactic for enhancing personally significant change in the treatment of adults and adolescents who stutter. *American Journal of Speech-Language Pathology, 21*(3), 264–277. https://doi.org/10.1044/1058-0360(2012/11-0068)

James, J. E., Ricciardelli, L. A., Rogers, P., & Hunter, C. E. (1989). A preliminary analysis of the ameliorative effects of time-out from speaking on stuttering. *Journal of Speech, Language, and Hearing Research, 32*(3), 604–610. https://doi.org/10.1044/jshr.3203.604

Johnson, G., Onslow, M., Carey, B., Jones, M., & Kefalianos, E. (2024). Lidcombe Program telehealth treatment for children 6–12 years of age: A Phase II trial. *Journal of Fluency Disorders, 80*, Article 106057. https://doi.org/10.1016/j.jfludis.2024.106057

Johnson, G., Onslow, M., Horton, S., & Kefalianos, E. (2023). Reduced stuttering for school-age children: A systematic review. *Journal of Fluency Disorders*. , Article 106015. https://doi.org/10.1016/j.jfludis.2023.106015

Jones, M., Onslow, M., Packman, A., & Gebski, V. (2006). Guidelines for statistical analysis of percentage of syllables stuttered data. *Journal of Speech, Language, and Hearing Research, 49*(4), 867–878. https://doi.org/10.1044/1092-4388(2002/019)

Kohmäscher, A., Primaßin, A., Heiler, S., Avelar, P. D. C., Franken, M. C., & Heim, S. (2023). Effectiveness of stuttering modification treatment in school-age children who stutter: A randomized clinical trial. *Journal of Speech, Language, and Hearing Research, 66*(11), 1–15. https://doi.org/10.1044/2023_JSLHR-23-00224

Koo, T. K., & Li, M. Y. (2016). A guideline for selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Laiho, A., Elovaara, H., Kaisamatti, K., Luhtalampi, K., Talaskivi, L., Pohja, S., Routamo-Jaatela, K., & Vuorio, E. (2022). Stuttering interventions for children, adolescents, and adults: A systematic review as a part of clinical guidelines. *Journal of Communication Disorders*. , Article 106242. https://doi.org/10.1016/j.jcomdis.2022.106242

Leósdóttir, B. (2014). Heildarmat á upplifun barna sem stama. Þýðing og staðfærsla OASES-S. [Overall assessment of the speaker's experience of stuttering (OASES-S): translation]. Master thesis. University of Iceland. ⟨http://hdl.handle.net/1946/18522⟩.

Lincoln, M., Onslow, M., Lewis, C., & Wilson, L. (1996). A clinical trial of an operant treatment for school-age children who stutter. *American Journal of Speech-Language Pathology, 5*(2), 73–85. https://doi.org/10.1044/1058-0360.0502.73

Mallick, R., Kathard, H., Thabane, L., & Pillay, M. (2021). A scoping review of the school-aged stuttering intervention literature. *Evidence-Based Communication Assessment and Intervention, 15*(4), 194–206. https://doi.org/10.1080/17489539.2021.1976964

Martin, R. R., Haroldson, S. K., & Triden, K. A. (1984). Stuttering and speech naturalness. *Journal of Speech and Hearing Disorders, 49*(1), 53–58. https://doi.org/10.1044/jshd.4901.53

Nippold, M. A., & Packman, A. (2012). Managing stuttering beyond the preschool years. *Language, Speech and Hearing Services in the Schools, 43*(3), 338–343. https://doi.org/10.1044/0161-1461(2012/12-0035)

Onslow, M., Webber, M., Harrison, E., Arnott, S., Bridgman, K., Carey, B., … Hearne, A. (2021). The Lidcombe Program treatment guide (Ver.1.3). Retrieved from ⟨https://www.uts.edu.au/sites/default/files/2021-04/Lidcombe%20Program%20Treatment%20Guide%202021%20v1.3%202021-04-27.pdf⟩.

Sjøstrand, Å., Kefalianos, E., Hofslundsengen, H., Guttormsen, L. S., Kirmess, M., Lervåg, A., Hume, C., & Naess, K. A. B. (2021). Non-pharmacological interventions for stuttering in children six years and younger. *Cochrane Database of Systematic Reviews*, (9)https://doi.org/10.1002/14651858.CD013489.pub2

Swift, M. C., & Langevin, M. (2024). A theory building critical realist evaluation of an integrated cognitive-behavioural fluency enhancing stuttering treatment for school-age children. Part 1: Development of a preliminary program theory from expert speech-language pathologist data. *Journal of Fluency Disorders, 82*, Article 106076.

Tichenor, S. E., Herring, C., & Yaruss, J. S. (2022). Understanding the speaker's experience of stuttering can improve stuttering therapy. *Topics in Language Disorders, 42*(1), 57–75. https://doi.org/10.1097/tld.0000000000000272

Tichenor, S. E., & Yaruss, J. S. (2021). Variability of stuttering: Behavior and impact. *American Journal of Speech-Language Pathology, 30*(1), 75–88. https://doi.org/10.1044/2020_AJSLP-20-00112

Van Riper, C. (1973). *The treatment of stuttering*. Prentice Hall.

Wagaman, J. R., Meltenberger, R. G., & Arndorfer, R. E. (1993). Analysis of a simplified treatment for stuttering in children. *Journal of Applied Behavior Analysis, 26*(1), 53–61. https://doi.org/10.1901/jaba.1993.26-53

World Health Organization. (2001). *International classification of functioning, disability, and health*. Author.

Yaruss, J. S. (2010). Assessing the quality of life in stuttering treatment outcomes research. *Journal of Fluency Disorders, 35*(3), 190–202. https://doi.org/10.1016/j.jfludis.2010.05.010

Yaruss, J. S., Coleman, C. E., & Quesal, R. W. (2012). Stuttering in school-age children: A comprehensive approach to treatment. *Language, Speech and Hearing Services in Schools, 43*, 536–548. https://doi.org/10.1044/0161-1461(2012/11-0044)

Yaruss, J. S., & Quesal, R. W. (2006). Overall assessment of the speaker's experience of stuttering (OASES): Documenting multiple outcomes in stuttering treatment. *Journal of Fluency Disorders, 31*(2), 90–115. https://doi.org/10.1016/j.jfludis.2006.02.002

Dr. Jóhanna Thelma Einarsdóttir is professor at the University of Iceland, in School of Education and School of Health. She is currently a chair of the speech and language pathology program at the University of Iceland. She has been employed as a Speech and language pathologist since 1987 specializing in stuttering since 1994. Her research has focused on stuttering, epidemiology of early stuttering measures and treatment of stuttering as well as general language acquisition, bilingual studies and developmental language disorders.

Manual for parents and children

Efficacy of a parent-managed response-contingent stuttering treatment program for older-aged children who stutter

Roger J. Ingham

Irene Seybold

Jóhanna T. Einarsdóttir

Background

Many research studies have demonstrated that a brief pause (time-out) following a moment of stuttering can result in a significant improvement in fluency for young children and adults who stutter. This project aims to investigate whether a structured program of time-out therapy for school-aged children can result in lasting improvements in fluency across a range of everyday situations. The program was first named TOHP or the Time-Out Home Program but was later renamed as STOC or Stuttering Treatment for Older Children. There is a great need to investigate systematically the efficacy of stuttering therapy for older school children in the search for efficient ways to increase fluency for the children. That's the purpose of this study of STOC.

You have the opportunity to participate in this project because your child responds positively to the time-out technique used in STOC (see Appendix). It is important that improvements in fluency are relevant in your child's daily life, rather than only within a therapy setting with their speech and language therapist. For this reason, the therapy involved in this research will take place within your child's everyday environment. As this research project aims to establish whether STOC is effective in treating stuttering, it is important that your child is not participating in any other therapy for stuttering while participating in this study.

The STOC treatment program

The STOC program is designed so that you, as parents, and speech and language pathologist train your child to self-administer the treatment method in order to produce more fluent speech. The format is designed so that your child will first learn to be completely fluent with natural-sounding speech during 10-minute morning and evening treatment sessions conducted with you at home (all sessions are video recorded). Next, the aim will be for your child to retain this fluent speech as treatment sessions become less frequent, changing gradually from daily to monthly.

This research project will be carried out over at least 2 years. We anticipate that you and your child will actively be working on the treatment program for at least a year of this time. The first stage will be collecting baseline measurements of your child's stuttering, and this will take either 3 or 6 months (Baseline). Your child will be randomly assigned to either a 3- or 6-month baseline period. The treatment starts after this baseline period. There will be two treatment periods with a two month break from treatment between these periods. The first treatment period (Treatment A) is 2 months in length and the second treatment period (Treatment B) is around 6 months in length. The duration of the Treatment B will vary according to your child's fluency. After Treatment B is complete, we will continue to evaluate your child's speech. This will occur once a month for the three months immediately after Treatment B is complete, and then again after 6 months (9 months after Treatment B was completed).

Monthly progress recordings

Every month during the entire project you will be asked to complete a number of tasks with your child which will be called monthly evaluations. These tasks will include you making recordings of your child speaking in different situations, both at home and beyond. This will allow us to evaluate your child's speech and fluency in a variety of situations before, during and after treatment. Monthly recordings will begin during the baseline period and continue until nine months after the treatments end.

Treatment periods

You and your child will both receive training in time-out therapy. You will video record every treatment session conducted at your home and send the recording to us after every session. All recordings will be watched be a clinician on the research team and you will receive feedback about how the treatment session conducted and a measurement of stuttering observed in the treatment session. Additional training on the treatment technique will be provided if/when necessary. Initially you will be in daily contact with the clinician as treatment sessions will be conducted daily. The frequency of the treatment sessions will gradually decrease to every 2nd day and to once every months, depending on changes in your child's fluency. The first treatment session recording should occur in the morning or first half of the day. In this session you, as the parent, directs the session by identifying stuttered syllable during the conversation and asking your child to pause speaking for a few seconds. The second treatment session should occur in the afternoon or evening. In this session your child will direct the session and needs to self-identify stuttered syllables and self-initiate pauses in speech following stuttered syllables.

Progress and relapse

It is normal to experience both periods of progress and relapse while undertaking treatment for stuttering. The treatment schedule is structured in a dynamic way, so that in periods of progress and fluency treatment sessions will occur less frequently. Conversely, in periods of relapse treatment sessions will occur more frequently according to the schedule. The research team will provide clear instructions about the frequency of your treatment sessions at every stage of the project. But remember, relapse is a normal, healthy part of the recovery process.

Below is more information about:

1. **The monthly assessment sessions**
2. **Treatment training**
3. **Treatment schedule**

1. MONTHLY ASSESSMENTS

As this is a research project, we want to be sure that we can attribute any changes or indeed, lack of changes, in fluency to STOC. It is therefore important to observe your child's speech throughout the treatment. Each month you will send us recordings of the child's speech. This will allow us to evaluate and compare his/her/their speech and fluency in a variety of situations before, during and following treatment. Each of these samples will be evaluated for: syllables spoken per minute, stuttering frequency, speech naturalness (a 1-9 rating each 60 seconds), and stuttering severity (a 0-10 rating at the end of the recording). The speaking situations are both within home and beyond home. These recordings begin when you agree to participate in the program and last throughout the project.

Within Home Speaking Tasks

You will record your child completing two tasks: (a) 3 minutes of reading aloud, and (b) 3 minutes of speaking in monologue.

Condition 1 (Home)	Time	Task
1	3 min	Reading
2	3 min	Monologue

Beyond-home Speaking Tasks

To go hand in hand with each within-home recordings the child must obtain at least three different recordings of their speech in routine speaking conditions. The first two speaking task i.e. conversing with a friend or relative and self- identified problematic speaking situation are to be recorded in situations outside of the home. The telephone call, however, can be made within the home situation. These additional tasks are suitable only if it can be recorded for 10-min once every month during the program. The SELF-SELECTED recording is to be made in a situation that your child considers to be problematic. This should be a speaking task that your child believe is particularly difficult but one in which your child would particularly like to be able to speak fluently. In other words, this situation should be one in which it will be possible to adequately gauge the effectiveness of the treatment program.

Condition 2 (Beyond Home)	Time	Task
1	10 min	conversing with a friend or relative
2	10 min	speaking on the telephone to a friend or relative
3	10 min	self-identified problematic speaking situation

Important notes regarding the recordings at home

These recordings are absolutely essential. They provide us with the most important data for evaluating the effects of the treatment program. Unless we have these recordings, we cannot proceed with the program. Please note the following concerning the recordings.

1. Please make separate recordings for each speaking task, one for reading, one for monologue etc.
2. Label the digital file with your name, the date of the recording, and the task e.g., (Mike Smith, Conversation with mom, 8.6.2014).
3. Once you have saved and renamed the file, please send it to your shared Google Drive account that you are using for this study.
4. You should make sure you are fully familiar with your recorder, iPad touch, and iPhone. Please ensure that the batteries have sufficient "life" before starting a recording.

5. Before you begin recording, please make a practice recording to make sure that you are recording your voice and your face clearly. Remember, we cannot assess your speech performance accurately unless we can hear and see both you and your child clearly.
6. You should try, as much as possible, to avoid noisy environments. Please do not record with a TV or radio playing in the background or in a moving car.
7. You should not record another speaker without their knowledge. In the case of telephone recordings, for example, you should ensure that only your child's speech is being recorded.
8. Please make sure that the speech samples you have recorded do not contain information that you would prefer NOT to be heard by a stranger. These recordings will be watched and evaluated by the researchers. If there is something in a recording you would prefer not to share, please rerecord the task. Although we generally pay little attention to what is being said in recordings, only to how it is said, you should be aware of this.

2. TREATMENT TRAINING

Parent and the Child Treatment Training

Training in STOC occurs after your child's baseline period has been completed. The training occurs at the start of the treatment period. First you, as the parent, will learn to deliver the time-out and then your child will learn to do the same. The baseline videos will be used for training you both on identifying stuttering occasions and to show how Time Out works in the context of therapy sessions.

In essence, training in STOC is implemented in the TO in the following way. When the child stutters during the speaking task (reading or conversation), the clinician immediately indicates "STOP" at which point the child should pause their speaking. After saying "STOP", the child and clinician remain silent for a 2-3 second pause, then the clinician simply says "Go on" and the child continues speaking (reading or conversation). The child does not have to repeat the stuttered word.

After observing this, the parent then takes over running the session and says "STOP" to implement the TO following them observing a stutter. Once the parent is reliable in doing this, the child practices pausing their speech without being prompted after the stutter. It is important that the clinician is satisfied that the parent and child are delivering the TO stimulus/event within 1-2 words after the perceived stutter. This is a critical phase of training and may require using an additional strategy until the child and the parent is as accurate as the clinician in delivering TO. The minimal level of accuracy in identifying a stuttered event and implementing a TO is 80% agreement.

3. TREATMENT

Treatment in general

It is essential that you exactly follow the procedure as you learned it in the training. Look carefully at the treatment schedule below. Each recording will be sent to the clinician through the app. You will receive feedback either through the app from your clinician or/and through email. Each time you think a step is complete you need to contact the clinician who determines if the rating is Pass or Fail. The minimal criteria for a "passed" morning and evening/afternoon recording are: (a) less than 1% of syllables stuttered, (b) ratings of 1-3 on the naturalness scale and 0 to 2 on the severity rating scale. You will be in regular contact with the clinician throughout the treatment and you and your child will get feedback on the treatment fidelity and the number of stuttered syllables. At first you will be in daily contact with your clinician, but then gradually there will be longer periods of time between contacts. The clinician or evaluator will rate the first 3 minutes of the child's speaking time of each recording (morning and evening).

Treatment periods

The first treatment period is for two months (*Treatment A*). How many sessions you will need to conduct at this time will vary according to how the treatment has been going. Some children will be practicing daily while others will need to practice less frequently. Then there is a break from treatment for two months. After two months the treatment starts again (*Treatment B*). This treatment block is for a minimum of 6 months. However, treatment may go for longer according to the individual need of your child. After this treatment block has concluded we will observe your child's speech for another 9 months.

Important notes regarding the treatment at home

1. The first treatment session should occur in the morning or the first half of the day. During this session, the parent administers TO after every stuttered syllable during the conversation.
2. The second treatment session should occur in the afternoon or evening. During this session child will administer self-initiated TOs.
3. Every treatment session must be recorded and the recording sent to the researchers. Send every therapy session at the end of the day. Identify the name of your child, the date of the recording and which recording is for a morning session and which is for an evening session.
4. Each treatment session should be 10 minutes in length (which ensures that your child is speaking for at least 3 minutes of the session).
5. It is important to note there should only be one recording for each session. There should be no pausing/erasing/re-record of the session to remove stuttering.
6. A clinician will evaluate the first 3 minutes of your child's speaking time of each recording (morning and evening). The recording will be analyzed for: syllables per minute, syllables stuttered, percent syllable stuttered (%SS), stuttering severity (0-10) and speech naturalness (1-9).

Treatment schedule

The schedule is outlined in the table below. Remember though that after two months treatment (Treatment A) there is no treatment for two months until the next treatment (Treatment B) begins.

The frequency of recordings will be based on a performance-contingent maintenance schedule (PCMS). Initially recordings will be made each day until three consecutive morning and afternoon recordings are judged to be a PASS. BOTH THE MORNING AND AFTERNOON RECORDINGS MUST ACHIEVE PASS CRITERIA FOR THE DAY SESSION TO BE JUDGED AS "PASS". The clinician assesses every recording during the initial phase on the every day level. Once your child has achieved almost no stuttering (ambiguous stuttering is allowed), SEV ≤ 2; NAT ≤ 3 at the *Every Day* level (see table below) on three occasions/days in a row and is judged to have passed, he/she/they can move to the *Every Other Day* level.

On the *every other day* level the clinician does not assess every recording. When the client is on the *every other day* level then the first two of three day recordings (e.g., *Every Other Day 1* and *Every Other Day 2* – see below) are scored by the parent (morning recording) and client (afternoon recording). Only the Every *Other Day 3* recordings are sent to the clinician who determines if the rating is Pass or Fail. Once three passes are achieved (2 from the client and parent's scores; 1 from the clinician) the client will progress to the next frequency level. The red bold text in the treatment schedule indicates that recordings from parent and client must be rated by the clinician – all others are recorded and rated by the client and parent. Their recordings and scores are to be forwarded immediately to the clinician.

Data Tracker

Data will be monitored and communicated with parent and client on an Excel sheet. The Excel sheet includes: Syllables per minute, syllables stuttered, percent syllable stuttered (%SS), naturalness rating and severity rating. This Excel sheet should also indicate the next step to be completed based on the treatment schedule below. Parents can follow the data tracker via their Google account.

Program Completion

Once your child has reached the end of the program, they are required to complete recordings (within-home and beyond-home recordings) once a month for 3 months and then after 6 months. These recordings are delivered by the procedure described above.

TREATMENT SCHEDULE Level/Frequency Attempted	**If Pass, progress to:**	**If Fail, retreat to:**
Every Day 1	Every Day 2	Every day 1R
Every Day 1R	Every Day 2	Every day 1R
Every Day 2	Every Day 3	Every day 1R
Every Day 3	**Every Other day 1**	Every day 1R
Every Other Day 1	**Every Other Day 2**	Every Other Day 1R
Every Other Day 1R	**Every Other Day 2**	Every Day 1
Every Other Day 2	**Every Other Day 3**	Every Other Day 1R
Every Other Day 3	**Every Third Day 1**	Every Other 1R
Every Third Day 1	Every Third Day 2	Every Third Day 1 R
Every Third Day 1R	Every Third Day 2	Every Other Day 1
Every Third Day 2	Every Third Day 3	Every Third Day 1 R
Every Third Day 3	**Every Fourth Day 1**	Every Third Day 1 R
Every Fourth Day 1	Every Fourth Day 2	Every Fourth Day 1R
Every Fourth Day 1R	Every Fourth Day 2	Every Third Day 1
Every Fourth Day 2	Every Fourth Day 3	Every Fourth Day 1R
Every Fourth Day 3	**Every Fifth Day 1**	Every Fourth Day 1R
Every Fifth Day 1	Every Fifth Day 2	Every Fifth Day 1R
Every Fifth Day 1R	Every Fifth Day 2	Every Fourth Day 1
Every Fifth Day 2	Every Fifth Day 3	Every Fifth Day 1R
Every Fifth Day 3	**Every Sixth Day 1**	Every Fifth Day 1R
Every Sixth Day 1	Every Sixth Day 2	Every Sixth Day 1R
Every Sixth Day 1R	Every Sixth Day 2	Every Fifth Day 1
Every Sixth Day 2	Every Sixth Day 3	Every Sixth Day 1R
Every Sixth Day 3	**Every Seventh Day 1**	Every Sixth Day 1R
Every Seventh Day 1	Every Seventh Day 2	Every Seventh Day 1R
Every Seventh Day 1R	Every Seventh Day 2	Every Sixth Day 1
Every Seventh Day 2	Every Seventh Day 3	Every Seventh Day 1R

Every Seventh Day 3	**Once Every 14 days 1**	Every Seventh Day 1R
Every 14 Days 1	Every 14 Days 2	Every 14 Days 1R
Every 14 Days 1R	Every 14 Days 2	Every Seventh Day 1
Every 14 Days 2	Every 14 Days 3	Every 14 Days 1R
Every 14 Days 3	**Once per month 1**	Every 14 Days 1R
	END OF PROGRAM	
Once per Month 1	Once per Month 2	Once per Month 1R
Once per Month 1R	Once per Month 2	Every 14 Days 1R
Once per Month 2	Once per Month 3	Once per Month 1R
Once per Month **3**	**Four Months check**	Once per Month 1R
	END OF MONITORING	

- R= Retry
- **Red = clinician scored**
- Black = client scored

Appendix

Testing Responsiveness to Time Out

To evaluate a child's responsiveness to TO as a means of decreasing stuttering behaviour, each child engaged in the following TO trial. The child reads aloud a written passage three time without use of TO. The child then repeats this, but this time the clinician delivers TO as the child reads, saying "STOP" within 1-2 words of a stutter occurring and requiring the child to pause from speaking for 2-3 seconds. Next the child's parent delivers the TO while the child is reading aloud a passage for 1 minute, three times. The child then repeats the three reading aloud tasks with no TO being used. This is then repeated with conversation as the speech context rather than reading a passage. Evaluation of both these contexts (reading aloud and conversation) with the child self-initiating TO.

TO was considered to be effective for the child is a 50% reduction in the percentage of syllables stuttered (%SS) compared to the baseline rate was observed in any of the contexts where TO was used.

Task	Time	Condition	Time-out
Oral Reading	1 minute	1A	NO
Oral Reading	1 minute	2A	NO
Oral Reading	1 minute	3A	NO
Oral Reading	**1 minute**	**1B**	**YES Clinician**
Oral Reading	**1 minute**	**2B**	**YES Clinician**
Oral Reading	**1 minute**	**3B**	**YES Clinician**
Oral Reading	**1 minute**	**4B**	**YES Parent/Child**
Oral Reading	**1 minute**	**5B**	**YES Parent/Child**
Oral Reading	**1 minute**	**6B**	**YES Parent/Child**
Oral Reading	1 minute	1A2	NO
Oral Reading	1 minute	2A2	NO
Oral Reading	1 minute	3A2	NO
Conversation	1 minute	1A	NO
Conversation	1 minute	2A	NO
Conversation	1 minute	3A	NO
Conversation	1 minute	1B	**YES Clinician**
Conversation	1 minute	2B	**YES Clinician**
Conversation	1 minute	3B	**YES Clinician**
Conversation	1 minute	4B	**YES Parent/Child**
Conversation	1 minute	5B	**YES Parent/Child**
Conversation	1 minute	6B	**YES Parent/Child**
Conversation	1 minute	1A2	NO
Conversation	1 minute	2A2	NO
Conversation	1 minute	3A2	NO

Paper IV

The Effect of Fidelity in a Pilot Study of Stuttering Intervention for Older Children

Íris Ö. Bergþórsdóttir[1], Jóhanna T. Einarsdóttir[1], Kathryn Crowe[1,2], Þorlákur Karlsson[3]
[1]University of Iceland, Iceland
[2]Charles Sturt University, Australia
[3]Brandr Research, Iceland

Author notes:

Íris Ö. Bergþórsdóttir: ORCiD 0000-0002-5121-3503

Jóhanna T. Einarsdóttir: ORCiD 0000-0001-8281-5331, Twitter @JohannaEinarsd1

Kathryn Crowe: ORCiD 0000-0003-3496-5129, Twitter @DrKateCrowe

Þorlákur Karlsson: ORCiD 0000-0002-2358-0402

Corresponding Author: Iris Ö. Bergþórsdóttur, School of Health Science, University of Iceland, Sæmundargötu 2, 101 Reykjavík. E-mail: iob@hi.is

Conflict of Interest: The authors have no conflicts of interest to declare.

Funding: This work was supported by the Icelandic Research Fund under grant 141029-052 awarded to J.T. Einarsdóttir.

Acknowledgments: Special thanks to Roger J. Ingham for his helpful contribution. The following clinicians and research assistants are acknowledged for their work on the study: Ása Birna Einarsdóttir, Berglind Bjarnadóttir, Bjarnfríður Leósdóttir, Guðrún Sigmundsdóttir, Karen Ingadóttir, Rósa Hauksdóttir, and Sigfús Helgi Kristinsson.

Disclosure of Interest: The authors report no conflict of interest.

Word Count: 9772 (including references)

Abstract

This paper evaluates the relation between fidelity of implementation (FOI) and children's outcomes in a study using the stuttering intervention for older children (STOC). Seven children aged 9–13 years received STOC, which consisted of parent- and child-managed sessions. Sessions were analysed for speech outcome measurements and four FOI components. Higher dosage was associated with less stuttering, talking more and lower severity ratings. Higher adherence was associated with less stuttering, more natural speech, and lower severity rating. FOI scores were higher for the children who finished STOC compared to children who did not. FOI contributes to understanding variation in outcomes.

Keywords: Fidelity, intervention, stuttering.

1.0 Introduction

The quest towards understanding not only *if* an intervention is effective but *why* it proves effective is a subject that evidence-based research has been focusing on more in recent years. Over time, there has been evolution in guidelines governing recommendations for the assessment and implementation of intervention fidelity in behavioural interventions, with the introduction of diverse frameworks for considering fidelity being proposed [1,2,3,4,5]. This construct, fidelity of implementation (FOI), can be defined as the measurement of whether the intervention was implemented as described during an intervention [6]. Measurement of FOI can simply be described as quantifying the gap between what was intended to happen during the intervention to what actually was done during the intervention.

It can be argued that researchers assessing intervention outcomes have for the longest time assumed that the interventions under study were executed competently, consistently, and accurately [7]. However, among the numerous variables recognized to influence intervention efficacy, FOI stands out as the most immediate and decisive factor [8]. The absence of consistent reporting of FOI in empirical articles describing behavioural intervention outcomes underscores a significant weakness that warrants attention in the field of behavioural science [9]. This absence of reporting is especially noticeable in research on interventions that take place in home settings [5,10]. The effectiveness and impact of an intervention are diminished when the intervention is not implemented in the manner it was originally designed and tested [11]. Failing to consider this aspect can greatly distort outcome interpretations, resulting in either overestimating or underestimating the true efficacy of the intervention. Without properly accounting for fidelity, researchers and clinicians are left with an incomplete picture of the intervention's overall impact. This may be one of the reasons why interventions that have well established effects in research settings have been shown to have diluted effects when delivered in real world settings [12]. Partnerships between speech-language pathologists (SLPs) and educators facilitate the development and application of strategies that enhance fluency in both academic tasks and social interactions in the school setting. By thoroughly analysing the child's speech patterns, environmental factors, and social dynamics, therapists can design targeted intervention plans to address the unique needs and challenges faced by each child [13,14,15]. Empathizing the importance of balancing protocols with clinical judgment and personalized care, rather than advocating for a one-size-fits-all approach.

Recent attention to FOI has meant that researchers' understanding of FOI is increasing and that it extends beyond simply reporting whether the correct dosage of intervention was received by participants, with new studies highlighting that FOI is more complex and multidimensional than previously considered [16].

Generally, FOI is defined as the level to which an intervention is implemented as planned [1,6,16,17,18,19]. The ability to interpret the effectiveness of interventions becomes higher when adequate fidelity measurements in research of interventions are made [20,21]. The essence of FOI is to methodologically measure the accuracy and consistency of implementations of interventions [2]. A core element to interpreting intervention effectiveness is to be able to evaluate and measure if the intervention described was carried out according to the research plan [6]. It is, therefore, important to consider both intended fidelity (as laid out in the research protocol) and achieved fidelity (as measured in the research conducted) [10,16,20]. Due to many different factors, the achieved fidelity is generally lower than the intended fidelity, and the fluctuation of how much the achieved fidelity varies from the intended fidelity can influence the intervention outcome [7,10]. The achieved fidelity can also vary greatly between participants within the same study. High FOI is linked to more robust and easily interpretable intervention outcomes, instilling greater confidence in delineating the underlying mechanism of change [22]). Incorrect conclusions can be drawn about effects, benefits, and comparisons in intervention studies when FOI is low, not considered, and/or not clearly reported [16,23].

A large range of aspects of fidelity have been used to measure FOI in previous research [7]. However, a unified definition of what FOI entails can be captured in four components: dosage, adherence, quality, and responsiveness. Dosage has been defined as the amount of intervention delivered, duration of each intervention session and the duration of the course of intervention [2,5,6,17,18,20,24]. Adherence has been defined as the extent to which all specified components of the intervention were accurately delivered, often measured by using independent observers, checklists, and automatic recording devices [2,5,6,17,18,20, 25]. Quality of intervention delivery encompasses the skill with which the intervention was delivered, the level of engagement in the intervention, and intervention protocols awareness [25,26,27]. Responsiveness of the participant to the intervention also considers the participant's engagement and participation and evaluating participant satisfaction by checklist or direct observation [2,3,5,17,18,24]. Together, these fidelity components function as adaptable guidelines applicable to a range of behavioural interventions.

1.1 The Importance of FOI in Intervention Research for Children Who Stutter (CWS)

Effective interventions to reduce stuttering are beneficial to CWS not only to support clear communication but to mitigate negative effects such as a heightened risk of social anxiety and depression [28]. While studies have shown the effectiveness of various therapeutic approaches for CWS [29,30], there can be many factors resulting in diminishing stuttering in CWS that are independent of the effect of an intervention. It has been documented that the recovery rate for CWS varies greatly across studies [31] and the high rate of spontaneous recovery of CWS is an example of one factor that makes determining whether an intervention was effective very complicated [32]. If many CWS experience spontaneous recovery in a study this might give a non-effective intervention a strong positive effect [33]. Conversely, an effective intervention may show a very small effect if few CWS within the sample spontaneously recover [33]. As various factors can contribute to a reduction in stuttering among CWS, such as spontaneous recovery, irrespective of the intervention, clinicians may find their confidence in the outcomes easily challenged [33]. As these factors aren't within the researcher's control, ensuring a clear and explicit measurement of FOI is crucial and enhances the comparability of outcomes across studies, thereby strengthen confidence in study findings.

To date, FOI measurements have not been systematically collected and reported in interventions for CWS [10], let alone been used to explicitly support the interpretation of intervention outcome effects. A review of 36 intervention studies that included a parent-implemented component for CWS (in addition to sessions implemented in a clinical setting) examined how information on FOI was presented in these studies, with reporting classified as evidence/direct measurement, based on a report, unclear, or not reported [10]. When considering if studies provided evidence that fidelity was examined for the parent-implemented sessions, evidence was only available in 19.4% of studies for dosage, 5.6% of studies for adherence, 8.3% of studies for quality, and responsiveness was never reported.

Results of the above review [10] of parent-implemented intervention for CWS showed that without the knowledge of how these interventions were implemented it is difficult to establish if an intervention was the cause of recovery or if other factors were at play, such as spontaneous recovery. A second challenge in research for CWS is the intervention is primarily conducted by the parents of the CWS in home-based sessions. Suggesting the fact that intervention is not being delivered by trained professionals, such as a speech-language pathologist (SLP) makes comprehensive measurement of fidelity even more important [10]. This is because measurement of only dosage will not capture differences in the adherence, quality, and responsiveness of delivery that may result from parents delivering intervention differently to how a professional would. None of the 36 studies examined/reported on all four FOI components for both clinical and home interventions.

1.2 The Current Study

Stuttering Treatment for Older Children (STOC) is a home-based intervention in which the child and the parent hold responsibility for identifying stuttering events [34]. The details of STOC and the pilot study are fully described in [34], with information relevant to the current study summarised here. STOC, like the Lidcombe Program [35] is built on the principles of Response Contingent Time Out (RCTO) which utilises contingency management to support fluent speech rather than altering speech patterns [14]. Intervention sessions were conducted twice a day. In the morning session the parent was responsible for identifying stuttered speech and prompting the child to pause speaking for 2–3 seconds. In the afternoon session the child was responsible for self-identifying stutters and pausing from speaking for 2–3 seconds. Sessions lasted 10 minutes each and were video recorded. In contrast to other RCTO approaches, STOC uses a performance-contingent maintenance schedule in which the frequency of intervention sessions is increased or decreased based on the occurrence or absence of stuttering events.

The purpose of the present study is to investigate FOI measures in detail using data from the STOC pilot study and to examine relations between FOI and children's outcomes. The study presents detailed FOI measures (dosage, adherence, quality, and responsiveness), the rationale for which are explained in detail in [10]. The FOI components in this study are defined as follows; *dosage*, referred to how often the intervention sessions were conducted; *adherence* was evaluated based on how precisely the intervention was delivered; *quality* was evaluated based on how well the intervention was delivered; and *responsiveness* was a judgement of how well the participants engaged in the intervention [2,5,6,10,17,18,20,24]. Few studies have used these FOI measurements in combination and with the purpose of examining variation in outcome measurements in general [5,25], but especially in intervention research related to CWS [10].

This paper aims to examine the relation between the FOI measures and children's outcomes using data from the STOC intervention study for CWS. The following research questions are addressed:

1. What is the relation between each FOI measurement (dosage, adherence, quality, and responsiveness) and each outcome measurement (%SS, SPM, naturalness, and severity) for children participating in the STOC pilot study?
2. Do differences exist in the FOI measurements between participants who completed the STOC intervention program and those who did not?

2.0 Method

2.1 Participants

Children were recruited to the study through the referral of a SLP or a special educational teacher – seven male children aged 9 to 13 years (*M* = 11;5, *SD* = 1;7). The criteria for inclusion of children in the STOC pilot study were as follows: (a) the child produced at least 3%SS in clinical settings either in oral reading or conversational speaking tasks; (b) the parents and the child were concerned about the child's stuttering; and (c) the parents and the child were willing to participate in the study. Exclusion criteria were the presence of a severe developmental language disorder and/or a developmental delay. The onset of the stuttering reported by the parents ranged from 2 to 10 years of age (*M* = 4;5, *SD* = 2;8). Participants' characteristics are presented in Table 1. Further detailed information about the participants is available in the report of STOC pilot study outcomes [34].

(Insert Table 1 here)

2.2 Procedure

2.2.1 STOC Intervention

Ethical approval was obtained from [blinded for review]. Both the parents and the children were trained to deliver the STOC intervention in a clinical setting before the start of the intervention period. Each intervention session lasted 10-minutes and was video-recorded using an iPod Touch via a custom-designed app. Videos were delivered to researchers for feedback after each intervention session was completed. Each session occurred twice daily, one in the morning and one in the afternoon. Parents managed the delivery of RCTO in the morning intervention sessions, prompting the child to pause from speaking when they identified a stuttering event. Following this, the child was required to pause in their speaking for 2-3 seconds. Children self-delivered RCTO in the afternoon sessions and were required to pause from speaking for 2-3 seconds each time they identified a stuttering event in their own speech. Parents received feedback after each session from a research SLP via email if difficulty with the implementation was observed or the recording quality was not adequate. The frequency of intervention sessions was dependent on participant performance during sessions, with sessions decreasing in frequency as children became more fluent, and vice versa. Of the seven participants, four were defined as having completed the intervention. The definition of completion was having completed the STOC period as well as the follow up period with complete data. The three participants who were defined as not having completed the intervention did not meet the performance-contingent criteria of the study and did not complete the second intervention phase. The full STOC protocol is available in [34].

2.2.2 Fidelity of Intervention (FOI) Measurements

FOI along four parameters was examined for both parent- and child-delivered STOC sessions. The parameters examined were dosage, adherence, quality, and responsiveness using the following definitions and procedures. To calculate dosage a record was kept of each day on which intervention sessions should have been delivered and every intervention session for which a video was delivered. Dosage was evaluated by comparing the proportion of planned/required sessions to the number of sessions that were actually delivered. Adherence was evaluated based on how accurately the intervention was implemented according to procedure outlined in the study protocol and taught to the parents and children. A sample of three minutes of consecutive speech was analysed from the recording of each session. This sample was taken to account for the fact that the conversations within session were constantly ongoing and ensured that consistency of the measurement context across both sessions for the same participant and across participants.

Adherence was evaluated by an SLP who counted (a) the accuracy of the parent asking the child to pause when a stuttering event occurred in parent-managed sessions, (b) the accuracy of the child pausing his speech when a stuttering event occurred in child-managed sessions, and (c) the number of stuttering events in the three-minute sample from each session. Adherence was calculated for parent managed sessions by dividing the number of said stops by the number of stuttering events evaluated by the SLP. For child managed sessions the number of observed pauses was divided by the number stuttering events evaluated by the SLP.

Quality was evaluated based on a judgement of whether the intervention was implemented with adequate skill. A single rating of quality was given to each intervention session based on the entire 10-minute video for each session. Quality was operationalised as a rating of the *flow* of conversation on a four-point ordinal scale: 1 = the conversation is flowing, and the child responds with full-length answers; 2 = the conversation is mostly flowing but with some exceptions; 3 = the conversation is mostly forced with some exceptions (e.g., the child may only be replying with single words or short phrases); and 4 = the child is not taking part in the conversation (mostly yes and no answers). Quality ratings were made by trained judges who were graduate SLP students after the completion of the STOC study.

Responsiveness was coded by rating the observed happiness of the child's participation in the intervention session on a three-point ordinal scale: 1 = very happy/conversation good; 2 = mildly happy/conversation fine; and 3 = not happy to participate in the session/conversation (see Appendix). The responsiveness of the child in each intervention was coded by trained judges who were graduate SLP students after the completion of the STOC study. In accordance with the STOC study protocol [34] all recordings were scored for FOI except when the schedule changed from every day to every other day, in which case recordings were scored for every third recording.

2.3 Outcome Measurements

To evaluate the effect of the intervention and to control for variability in speech performance, before, during, and after intervention, monthly video recordings were collected. The Stuttering Measurement System program [36] was used to calculate all outcome measurements (%SS, SPM, naturalness, and severity). Measurements were completed by seven judges, two of which were qualified SLP and five of which were trained SLP students. Each judge got 15–20-hour training in accurately measuring stuttering through the program. The collection of outcome measurements was from May 2014 to June 2017. In total 653 monthly recordings were collected from the participants, almost 80 hours of recordings. Evaluation of %SS was done by dividing the number of stuttering occurrences by the number of syllables spoken. SPM was measured by counting the number of syllables spoken divided by three, which was the number of minutes of the speech considered from the session. Naturalness was evaluated by a 9-point ordinal scale (1 = highly natural sounding, 9 = highly unnatural sounding) [37]. Similarly, severity was evaluated using an 11-point ordinal scale (0 = no stuttering, 10 = severe stuttering) described in [38].

2.4 Reliability

Reliability was examined for data related to outcome measures by three SLP graduate students who were not involved in the original scoring. All were trained using the stuttering measurement system for 15–20 hours. They independently measured %SS and SPM and made naturalness and severity ratings for 116 (7.5%) of the 1551 (*Range* = 148–350, *M* = 222, *SD* = 81) intervention sessions. Reliability was calculated with intraclass correlation coefficients which were high, .92 for %SS, .87 for SPM, .78 for naturalness and for severity. Reliability of fidelity measurements was also calculated. As dosage was measured by comparing how many intervention sessions occurred compared to how many should have occurred, the first author recalculated both values for each child and found these were 100% accurate. For adherence, two judges independently rated 1471 (100%) recordings, with differences discussed until consensus was reached. For quality and responsiveness 1409 (95.8%) recordings were rated by two independent judges. The first author rated 161 (10.4%) of the recordings, including an equal number from each judge to evaluate the interrater reliability. Point-by-point reliability was 95.3%.

2.5 Data Analysis

Data analysis was conducted in Excel and Statistical Package for the Social Sciences (SPSS). When the recording of an intervention session was not received by the researchers, for whatever reason, it was treated as if the intervention session had not occurred (i.e., no dose for that day). This was applied regardless of whether the parent said that the intervention had occurred or not. No corrections for missing data were made in analyses. While ratings of stuttering outcomes and adherence were made close to the time that recordings were received by researchers, ratings for quality and responsiveness were made several years after the recordings were collected. During this time some of these recordings had become unusable, for example several files were corrupted and not usable. Consequently 4.2% recordings were missing from analysis for quality and responsiveness. To answer research question one, Pearson Product-Moment correlations were performed between each FOI variable (dosage, adherence, quality, responsiveness) and each outcome measurement (%SS, SPM, naturalness, severity). In cases where correlation was based on repeated measures within subjects yielding autocorrelation, the coefficients were corrected employing multiple regression [39]. Correlation strength was interpreted as .0–.3 weak, .3–.7 moderate and .7–1.0 strong [40]. Significance was set at $p < .05$ for two tail comparison, however, as the number of correlation coefficients was quite high, procedure of controlling false discovery rate (FDR) was employed [41]. To answer research question two, analysis of mean differences employing confident intervals (CI) was conducted.

3. Results

3.1 Description of Fidelity

The results for the fidelity of intervention (FOI) measurements are presented in Table 2. Overall mean dosage across parent- and child-managed sessions was 83.2% (not in the table) (*Range* = 58.5–97.7%), with 1551 intervention sessions delivered. There was some variation in the dosage achieved between the parent-managed (*M* = 85.9%) and the child-managed (*M* = 80.5%) intervention sessions. The overall mean adherence across parent- and child-managed sessions was 58.8% (not in the table) (*Range* = 12.8-80.9%). There was some variation in the adherence between the parent-managed (*M* = 64.6%) and the child-managed (*M* = 53.1%) intervention sessions. The overall mean quality across parent- and child-managed sessions (1 being good quality and 4 being poor) of the intervention delivery was 1.7 (not in the table) (*Range* = 1.0–2.2). There was little variation between the parent-managed (*M* = 1.7) and the child-managed (*M* = 1.6) intervention sessions. For responsiveness (1 being a good conversation and 3 being a poor) the overall mean across parent- and child-managed sessions was 1.1 (not in the table) (*Range* = 1.0–1.4). There was no variation between the parent-managed (*M* = 1.1) and the child-managed (*M* = 1.1) intervention sessions.

(Insert Table 2 here)

3.2 FOI and Outcome Measures

Pearson Product-Moment correlation coefficients were calculated to evaluate the impact of each FOI on each outcome measure. No significant coefficient was found between dosage and any outcome measure (see Table 3). However, 7 of the 12 comparisons were of moderate strength (> .29), but low N resulted in none being significant. Negative correlations were identified with %SS, indicating that higher dosage was associated with less stuttering and positive correlations were identified with SPM, meaning that higher dosage was associated with talking more. For naturalness and severity, negative correlation indicated that speech was more natural, and stuttering was less severe. All correlation coefficients between dosage and severity were moderately strong, whereas all coefficients with naturalness were practically zero. In child-managed intervention sessions, %SS, SPM, and severity were all moderately correlated.

(Insert Table 3 here)

In Table 4 correlation coefficients between adherence to the intervention and the four outcome measures are displayed for both parent- and child-managed intervention sessions. Of the parent-managed sessions, 17 of the 28 correlation coefficients for the participants separately were significant and 16 of the 28 for child-managed sessions. Most coefficients were weak, with only 11 of the 28 coefficients being of moderate strength for parent-managed sessions and 12 of the 28 coefficients for child-managed sessions and one was strong. Negative coefficients were identified with %SS, indicating that higher adherence was associated with less stuttering and negative coefficients were identified with SPM, indicating that higher adherence was associated with less speech, although only two coefficients were significant. For naturalness and severity, negative coefficients indicated that speech was more natural with more adherence, and stuttering was less severe. Most correlation coefficients between adherence and %SS were moderately strong whereas most coefficients between adherence and SPM were weak, in both parent- and child-managed intervention sessions. In child-managed intervention sessions three of the correlation coefficients between adherence and naturalness were of moderate strength and one showed strong correlation.

(Insert Table 4 here)

In Table 5 one can see correlation coefficients for both parent- and child-managed intervention sessions between quality, as a rating of the flow of conversation, and the four outcome measurements. Of the parent-managed sessions, 15 of the 28 correlation coefficients for the participants separately were significant and 9 of the 28 for child-managed sessions. Most of the correlations were weak with 8 of the 28 coefficients being of moderate strength, and one of strong coefficient for the parent-managed sessions and 8 of the 28 coefficients were of moderate strength for the child-managed sessions. Negative correlations were identified with %SS indicating that higher quality was associated with less stuttering, however, significant only in 3 cases of 14, and negative correlations were identified with SPM, indicating that higher quality was associated with talking more with most coefficients significant. For naturalness and severity, negative correlations indicated that the flow of the conversation was related to more natural sounding or less severe stuttering with 10 out of 28 coefficients significant.

Most correlation coefficients between quality and SPM were moderately strong (8 out of 14) and one strong, whereas most correlations between quality and %SS were weak (9 out of 14). Most correlation coefficients between quality and naturalness were weak (8 of 14), as well as between quality and severity (9 out of 14).

(Insert Table 5 here)

Table 6 displays correlation coefficients for responsiveness of the child in conversation to the four outcome measurements for both parent- and child-managed intervention sessions. Of the parent-managed sessions, 11 of the 28 coefficients for the participants separately were significant and only 6 of the 28 for child-managed sessions. Most coefficients were weak with only 3 of the 28 coefficients being of moderate strength for parent-managed sessions and 1 of the 28 coefficients were of moderate strength for child-managed sessions. Three significant negative correlation coefficients between responsiveness and %SS were identified, indicating that higher responsiveness was associated with less stuttering in these cases. Significant negative coefficients were also identified with SPM in half of the cases, indicating that higher responsiveness was associated with talking more. For naturalness and severity, both positive negative correlation coefficients were identified, depending on the participant. A positive relation means that high responsiveness, i.e. happy and good conversation, goes with more natural sounding speech and less severity of stuttering. The coefficients were all weak except for one and about half of them were .10 or lower, indicating that responsiveness was not related with naturalness and severity for most participants.

(Insert Table 6 here)

3.3 FOI and Intervention Completion

Fidelity measures were compared between the four children who completed the STOC intervention (P1, P3, P4, P5) and the three children who did not complete the intervention (P2, P6, P7). The dosage was higher for participants who completed the intervention, both in parent-managed (*M* = 92.7%) and child-managed (*M* = 87.6%) intervention sessions, compared to the group that did not complete the intervention (see Table 7). Parent-managed intervention sessions for the group that completed the intervention had the highest dosage overall. The overall dosage for the group that did complete the intervention was higher (*M* = 90.2%) than the group that did not complete the intervention (*M* = 73.9%) (not in the Table 7).

(Insert Table 7 here)

To evaluate differences in adherence, responsiveness, and quality for these two groups, analysis of mean differences employing CIs was conducted (see Table 8). For participants who completed the intervention, the adherence was higher, both for parent-managed (*M* = 74.0%) and child-managed (*M* = 66.8%) intervention sessions, than the group that did not complete the intervention (56.2% and 41.7% respectively). The lack of overlap in the CIs indicated that this difference between those who completed and did not complete is unlikely to be coincidental. For participants who completed the intervention, the quality had a better score, both for parent-managed (*M* = 1.46) and child-managed (*M* = 1.38) intervention sessions, than the group that did not complete the intervention (1.87 and 1.75 respectively). The lack of overlap in the CIs indicated that this difference between those who completed and did not complete is also unlikely to be coincidental. In regard to responsiveness, an almost complete overlap in the CIs indicates no difference between the groups, i.e, no significant difference between the two groups.

(Insert Table 8 here)

Discussion

In this study, we have assessed the relation between four FOI measurements (dosage, adherence, quality, and responsiveness) and four outcome measurements (%SS, SPM, naturalness, and severity) in a pilot study of an intervention for CWS. The findings suggest that among the four FOI components, dosage and adherence exert the most influence on reducing stuttering frequency as well as contributing to less severe stuttering. Increased adherence correlated with reduced stuttering, suggesting that effective utilization of the intervention led to decreased stuttering frequency. However, adherence did not appear to increase the overall amount of speech produced by the children. Better quality and responsiveness neither reduced stuttering frequency nor established more naturalness and less severity, it did, however, enhance the flow of parent-child conversations. Dosage and adherence were higher in the parent-managed intervention sessions in comparison to the child-managed sessions. When comparing the FOI measures for participants who completed the intervention and who did not complete interesting patterns were observed. FOI measurements were always better for the participants who completed the intervention compared to those who did not, suggesting that the intervention is more effective when the protocols of the intervention design are met. Using the OASES-S evaluation tool [42], results indicated that the four participants who successfully completed the intervention demonstrated notable enhancements in their fluency and overall quality of life [34].

The results of this study add to the literature on fidelity implementation and the importance of establishing measurements of FOI in intervention studies [17,25,43,44]. Results from the present study show that dosage plays a big role for decreased stuttering, increased speech and less severity, while dosage does not show a significant correlation with the outcomes due to the low sample size, it exhibits a moderate correlation in 7 out of 12 cases. This result is not surprising as dosage has long been considered an important aspect of FOI across interventions for diverse purposes (Drug abuse prevention: [18]; Early communication intervention: [5]; Psychotherapy: [25]; Preventive interventions: [17]. Most research on stuttering report on dosage in some form [10]. Results from studies on parent verbal contingencies interventions, such as the Lidcombe Program [45], indicate a link between correct usage of intervention and decreased stuttering severity, in home-based settings. In a study on syllable-timed speech there was reporting of reduction in beyond-clinic stuttering, associated with correct introduction of intervention [46]. This relation between adherence and decrease stuttering severity is also seen in the results of the present study.

However, unlike the present study, previous studies of interventions for CWS fail to evaluate how correct intervention effects outcomes [10].

Higher levels of quality and responsiveness enhance the flow of parent-child conversations, fostering engagement and encouraged positive participation. In the present study, quality was a measurement of how well the conversation was flowing, as without a conversation there can be no intervention. When studies of CWS report data on the quality of intervention, this is often not done in a systematic way. For example, a study may state that parents reported that some instructions of the procedure were difficult to follow, without explanation on what in the instructions was hard to follow, how or if this was corrected and how this might have affected the results. However, even when quality as FOI is considered in any form, studies do not link these data directly to outcomes [47,48]. For example, [48] reported that parents did not provide sufficient structure in the speaking session with the child, interrupted the child when speaking, and in some cases prevented the child from speaking when they felt it was their turn to speak. While [48], among others, have reported on quality in intervention as an incidental and anecdotal sidenote, this pilot of the STOC intervention measured quality systematically and examined the impact of quality on outcomes. Similarly, reports of responsiveness as an aspect of FOI are severely lacking in stuttering intervention research, both for clinical settings and at home. Where this has been addressed in prior research, as by [49], this was through parent and/or child completed questionnaires, asking about perceived changes in stuttering and the acceptability of the intervention. However, in contrast to the present study, these measurements were not used to evaluate how the intervention was being administered or to analyse the impact of responsiveness on outcomes.

In order to participate in the STOC intervention pilot study, each parent-child dyad needed to spend at least 20 minutes daily in a one-to-one conversation. Research on work-family balance [50], show that time spent with family is important but often a challenge with the demands of modern lifestyles. The amount of time the intervention required each day may have presented some challenges for participants, but also created a space in which not only intervention was occurring, by child-parent relationships could be fostered. This positive experience may render the experience of intervention as painless and enjoyable, potentially motivating both parties. In a study utilising the Lidcombe program in early intervention, it was noted that mothers often struggled to incorporate the intervention sessions into their schedules and occasionally forgot to conduct sessions [51]. Similarly, it was reported that parents said that they had difficulties meeting the research requirements, but they were required to attend weekly clinic visits, provide research data weekly as well as they could not take breaks from the treatment, due to other day-to-day life responsibilities, such as the presence of younger children [45]. In the present study, factors such as this might have contributed to the high proportion of participants who did not complete the intervention.

Recommendations for Researchers

Neglecting to assess FOI in intervention research may result in an inaccurate link between intervention and outcome, jeopardizing the accuracy and generalizability of findings. In the absence of FOI data, researchers might erroneously deem an intervention ineffective when in fact the intervention's true efficacy could be substantial if it had been executed as originally intended [16]. Potentially, it could also provide evidence indicating that a reduced level of intervention is satisfactory.

Enhanced FOI data can offer deeper insight into the outcome measures influenced by intervention as the assumed contribution of the intervention may not align with its actual impact. A positive association between verbal contingencies and treatment progress is what is expected in the Lidcombe Program, yet results suggest that this might not be the case as findings reported that more verbal contingencies for stuttering was associated with a longer time to complete treatment [48]. Findings such as these raise questions about other published studies, especially in regards to the effectiveness of an intervention and what components of the intervention are vital to its success. Thus, it is evident that collection of FOI data at all stages of intervention research is imperative for a comprehensive evaluation of results, but this is also challenging for researchers. Some studies [52] suggest that better treatment outcomes are not necessarily related to strict implementation protocols, but FOI measurements can highlight the essential components of effective interventions, avoiding unnecessary steps that do not contribute to therapeutic success. Firstly, a lack of clarity over how FOI should be measured can impede researchers' measurement of FOI. There are large inconsistencies in definitions, measurements and analyses across fields [53], although consensus is gradually developing towards agreement on the core components of FOI [7,17,25,43,44,54,55]. Secondly, experts in FOI articulate the importance of research design, implementation protocols and output measures that consider FOI with relative clarity and consistency. However, there is a lack of agreement regarding how FOI can best be monitored and measured, which is essential to assessing the closeness between intended and achieved fidelity [5,10,25]. The results of the present study show that dosage and adherence were the FOI components that had the most effect on outcomes, which are the same components that have been most often reported in prior studies of parent-implemented interventions for stuttering, although not linked to the outcomes of these studies [10]. The present study also found that quality and responsiveness added little to explanations of variance in outcomes, which may be attributed to these constructs not being operationalised well in this study, the measurements used of quality and responsiveness not adequately capturing these constructs, or a lack of effect of these constructs on outcomes. Finally, it has been noted that researchers face logistical challenges in examining FOI in their research [56]. Some of the obstacles they raised included resource constraints, difficulties in designing and analysing measures, and challenges in training clinicians to fulfil these requirements. While the challenges that researchers face cannot be ignored, if efforts are not taken to measure and maximize FOI, study findings will lack validity and reliability [16] and could constitute waste of precious resources, such as research funding and the time of both participants and researchers. Fidelity measurements are, therefore, a highly important aspect to evaluate the outcome of intervention studies. Potential avenues for future research on FOI, that would better inform studies of interventions for CWS encompass a thorough examination of current fidelity frameworks and reporting guidelines for behavioural interventions, as well as guidance on optimal strategies for integrating fidelity data in intervention outcome research. We also propose that future research should explore the interplay between FOI and therapist flexibility, investigating how adherence to key components can coexist with personalized treatment strategies. This balanced approach may offer a more comprehensive understanding of the factors driving successful therapeutic outcomes, ultimately enhancing the quality of care provided to clients.

Recommendations for SLP’s

Clinicians need to carefully examine the measurement and reporting of FOI when reading research and deciding if research evidence supporting an intervention is sufficient to inform their evidence base for practice. Thorough reporting of FOI in published research also provides clinicians clear information on what parameters are crucial for implementing the intervention effectively to obtain the desired outcome. Then, clinicians would be able to implement the intervention as it has been found to be effective within their clinical practice. This also helps clinicians to engage in evidence-based discussion about intervention options with CWS and their families as clinicians can present the possible outcomes and implementation difficulties of different intervention [51] and support parents in making informed decisions about intervention options.

Understanding the strength of associations between specific FOI components and outcomes offers guidance on prioritizing elements during the implementation of interventions. For instance, if dosage and adherence emerge as key factor for child outcomes while responsiveness does not, practitioners may emphasize collaborating with families to enhance both the frequency and intervention accuracy of their interventions with their children. Clinicians need to monitor the FOI of intervention with each child/family to ensure that the intervention is being implemented with fidelity, supporting the child/family to achieve the required FOI, and monitoring if the expected outcomes are being achieved given the level of FOI that is present. If FOI is good but outcomes are not achieved, this provides evidence that this intervention may not be effective for this child and an alternative approach should be trialled. Without this FOI data, a clinician cannot know whether the approach is not working because it is inappropriate for that child or if it is because it has not been administered in the way it needs to be in order to be effective. Implementing interventions in clinical and home-based settings that lack fidelity also constitute a waste of resources, which includes the time of the clinician and child/family as well as public and/or private funding for clinical services. Clinicians often manage large caseloads, operating under constraints of limited time and resources. When interventions yield suboptimal results due to poor fidelity—whether from inconsistent protocol adherence or poorly executed home practices—clinicians are forced to invest additional time without seeing proportional progress in the child's development. This not only hampers the clinician's capacity to assist other patients but also adds undue pressure to their workload by having to address issues that higher fidelity could mitigate.

Limitations of This Study

This study has a number of limitations. First and foremost is the small sample size. Only seven children were eligible to participate, and only four of these participants completed the intervention. With a larger sample, it would have been possible to evaluate with more certainty differences between the group that completed and those that did not complete the intervention. This small sample might not correctly represent a broader group of CWS and restrict the ability to generalize findings to a broader population. The high rate of children who failed to complete the intervention can make it challenging to establish a clear cause-and-effect relationship, as the lack of completion may be influenced by various confounding factors. A larger sample of CWS might also give greater clarity to the relation between each FOI component and each outcome.

A second limitation comes from the interaction between the FOI components adherence and quality. Parents reported during the study that they were hesitant to say “stop” during a conversation when their child had started to speak more fluently or when subtle stuttering events occurred. Parents reported that they were happy that their child’s speech was becoming more fluent and natural and that they did not want to discourage the child by identifying the stuttering event. Thus, as quality increased, because the quality of the interaction in the intervention was improving as the child was speaking more, adherence would decrease as parents were not identifying all stuttering events. This interaction could impact the interpretation of results of the relation of these FOI measures to outcome measures. Third, when the stuttering events became less severe, the parents had difficulties recognizing more subtle stuttering events, which decreased adherence. Not identifying these subtle stuttering events decreases adherence while the severity of the stutter also decreases. This would lead to a reduction in the strength of correlations between adherence and outcome measure, making interpretation of the relation between adherence and outcomes difficult.

Fourth, the subjective nature of quality and responsiveness measurements presented challenges, and it was noticed that their evaluation appeared closely aligned in some cases. Quality is a measurement on the flow of the conversation. It was clear in many of the intervention videos that holding a 10-minute conversation twice a day was not easy for either parents or children. This was especially difficult as the interaction needed to go beyond yes/no questioning and to a conversation that both parent and child were engaged in. Responsiveness measured the engagement of the participations and could be affected by intervention fatigue, as well as sometimes the children were not in the mood for a conversation, regardless of the intervention.

Conclusion

Measuring implementation fidelity is important to help researchers understand the extent to which participant outcomes can be attributed to the intervention itself rather than variations in how it was implemented. The results of this study show that children who completed STOC had higher ratings on FOI measurements than those who did not complete the intervention. Dosage and adherence were associated with less stuttering, more naturalness, and less severity, while quality and responsiveness were associated with the child speaking more. The results show that when STOC was implemented as intended the intervention was more effective.

Acknowledgments: [remove for review]

Declaration of Interest: The authors report no conflict of interest.

References

[1] An, M., Dusing, S. C., Harbourne, R. T., Sheridan, S. M., & START-Play Consortium. (2020). What really works in intervention? Using fidelity measures to support optimal outcomes. *Physical Therapy, 100*(5), 757–765. https://doi.org/10.1093/ptj/pzaa006

[2] Bellg, A. J., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D. S., Ory, M., Ogedegbe, G., Orwig, D., Ernst, D., Czajkowski. S., & Treatment Fidelity Workgroup of the NIH Behavior Change Consortium. (2004). Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the NIH behavior change Consortium. *Health Psychology*, *23*(5), 443–451. https://doi.org/10.1037/0278-6133.23.5.443

[3] Century, J., Rudnick, M., & Freeman, C. (2010). A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Evaluation, 31*(*2*), 199–218. https://doi.org/10.1177/1098214010366173

[4] Gould, L. F., Dariotis, J. K., Greenberg, M. T., & Mendelson, T. (2016). Assessing fidelity of implementation (FOI) for school-based mindfulness and yoga interventions: A systematic review. *Mindfulness*, *7(1),* 5–33. https://doi.org/10.1007/s12671-015-0395-6.

[5] Lieberman-Betz, R.G. (2015). A systematic review of fidelity of implementation in parent-mediated early communication intervention. *Topics in Early Childhood Special Education, 35(1)*, 15–27. https://doi.org/10.1177/0271121414557282

[6] Perepletchikova, F., & Kazdin, A. E. (2005). Treatment integrity and therapeutic change: Issues and research recommendations. *Clinical Psychology: Science and Practice*, *12*(4), 365–383. https://doi.org/10.1093/clipsy.bpi045 https://doi.org/10.1037/0022-006X.75.6.829

[7] Hagermoser Sanetti, L. M., & Kratochwill, T. R. (2014). *Treatment integrity: A foundation for evidence-based practice in applied psychology*. American Psychological Association.

[8] Noell, G. H. (2008). Research examining the relationships among consultation process, treatment integrity, and outcomes. In W.P. Erchul & S. M. Sheridan (Eds.), *Handbook of research in school consultation: Empirical foundations for the field* (pp. 315–334). Erlbaum.

[9] Kechter, A., Amaro, H., & Black, D.S. (2019). Reporting of treatment fidelity in mindfulness-based intervention trials: A review and new tool using NIH behavior change consortium guidelines. *Mindfulness, 10,* 215–233. https://doi.org/10.1007/s12671-018-0974-4

[10] Bergþórsdóttir, Í. Ö., Crowe, K., & Einarsdóttir, J. T. (2022). Implementation fidelity in parent-implemented interventions for stuttering. *Clinical Linguistics & Phonetics*, *36*(10), 904–927. https://doi.org/10.1080/02699206.2021.1965659

[11] Walrath, C. M., Sheehan, A. K., Holden, E. W., Hernandez, M., & Blau, G. M. (2006). Evidence-based treatments in the field: A brief report on provider knowledge, implementation, and practise. *Journal of Behavioral Health Services & Research, 33*, 244–253. https://doi.org/10.1007/s11414-005-9008-9

[12] Feldman, M. B., Silapaswan, A., Schaefer, N., & Schermele, D. (2014). Is there life after Debi? Examining health behavior maintenance in the diffusion of effective behavioral interventions initiative. *American Journal of Community Psychology, 53*(3–4)*,* 286–313. https://doi.org/10.1007/s10464-014-9629-3

[13] Blood, G. W., & Blood, I. M. (2004). Bullying in adolescents who stutter: Communicative experience and self-esteem. *Contemporary Issues in Communication Sciences and Disorders, 31*, 69–79.

[14] Bothe, A. K., Davido, J. H., Bramlett, R. E., & Ingham, R. J. (2006). Stuttering treatment research. 1970–2005: I. Systematic review incorporating trial quality assessment of behavioural, cognitive, and related approaches. *American Journal of Speech-Language Pathology 15(4),* 321–341. https://doi.org/10.1044/1058-0360(2006/031)

[15] Guitar, B. (2013). Stuttering: An integrated approach to its nature and treatment (4th ed.). *Lippincott Williams & Wilkins*.

[16] Sanetti, L. M., Cook, B. G., & Cook, L. (2021). Treatment fidelity: What it is and why it matters. *Learning Disabilities, Research & Practice*, 36, 5–11. https://doi.org/10.1111/ldrp.12238

[17] Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control. *Clinical Psychology Review*, *18*(1), 23–45. https://doi.org/10.1016/S0272-7358(97)00043-3

[18] Dusenbury, L., Brannigan, R., Falco, F., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implementations for drug abuse prevention in school settings. *Health Education Research, 18*(2)*,* 237–256. https://doi.org/10.1093/her/18.2.237

[19] Prowse, P. D., Nagel, T., Meadows, G. N., & Enticott, J. C. (2015). Treatment fidelity over the last decade in psychosocial clinical trials outcome studies: A systematic review. *Journal of Psychiatry, 18*(2), 1000258. https://doi.org/10.4172/Psychiatry.1000258

[20] Kaderavek, J. N., & Justice, L. M. (2010). Fidelity: An essential component of evidence-based practice in speech-language pathology. *American Journal of Speech-Language Pathology*, *19,* 369–379. https://doi.org/10.1044/1058-0360(2010/09-0097)

[21] McCormack, J., Baker, E., Masso, S., Crowe, K., McLeod, S., Wren, Y., & Roulstone,

S. E. (2017). Implementation fidelity of a computer-assisted intervention for children with speech sound disorders. *International Journal of Speech-Language Pathology, 19*, 265–276. https://doi.org/10.1080/17549507.2017.1293160

[22] Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science, 2*(40), 1-9. https://doi.org/10.1186/1748-5908-2-40

[23] Bergþórsdóttir, I. Ö., & Ingham, R. J. (2016). Putting the cart before the horse: A cost effectiveness analysis of treatments for stuttering in young children requires evidence that the treatments analyzed were effective. *Journal of Communication Disorders, 65*, 65–67. https://doi.org/10.1016/j.jcomdis.2016.04.006

[24] Slaughter, S. E., Hill, J. N., & Snelgrove-Clark, E. (2015). What is the extent and quality of documentation and reporting of fidelity to implementation strategies: A scoping review. *Implementation Science*, *10*(1-12). https://doi.org/10.1186/s13012-015-0320-3

[25] Perepletchikova, F., Treat, T. A., & Kazdin, A. E. (2007). Treatment integrity in psychotherapy research: Analysis of the studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology, 75*(6), 829–841.

[26] Gearing, R. E., El-Bassel, N., Ghesquiere, A., Badwin, S., Gillies, J., & Ngeow, E. (2011). Major ingredients of fidelity: A review and scientific guide to improving quality of intervention research implementation. *Clinical Psychology Review, 31*(1), 79–88. https://doi.org/10.1016/j.cpr.2010.09.007

[27] Santacroce, S. J., Maccarelli, L. M., & Grey, M. (2004). Intervention fidelity. *Nursing Research, 53*(1), 63–66. http://doi.org/10.1097/00006199-200401000-00010

[28] Hofslundsengen, H., Kirmess, M., Guttormsen, L. S., Næss, K.A.B., & Kefalianos, E. (2022) Systematic review of implementation quality of non-pharmacological stuttering

intervention trials for children and adolescents. *Journal of Fluency Disorders*, *71,* 105884. https://doi.org/10.1016/j.jfludis.2021.105884

[29] Andrews, C., O'Brian, S., Onslow, M., Packman, A., & Menzies, R. (2016). Phase II trial of a syllable-timed speech treatment for school-aged children who stutter. *Journal of Fluency Disorders, 48*, 44–55. https://doi.org/10.1016/j.jfludis.2016.06.001

[30] de Sonneville-Koedoot, C., Stolk, E., Rietveld, T., & Franken, M-C. (2015). Direct versus indirect treatment for preschool children who stutter: The RESTART randomized trial. *PLoSONE, 10*(7). https://doi.org/10.1371/journal.pone.0133758

[31] Einarsdóttir, J. T., Crowe, K., Kristinsson, S. H., & Másdóttir, T. (2020). The recovery rate of early stuttering. *Journal of Fluency Disorders*, *64*, 105764. https://doi.org/10.1016/j.jfludis.2020.105764

[32] Yairi, E., Ambrose, N., & Grinager, N. (2005). *Early childhood stuttering*. Pro-Ed.

[33] Ingham, J. C., & Riley, G. (1998). Guidelines for documentation of treatment efficacy for young children who stutter. *Journal of Speech, Language, and Hearing Research*, *41*, 753–770. https://doi.org/10.1044/jslhr.4104.753

[34] Einarsdóttir, J. T., Karlsson, T., Bergþórsdóttir, Í. Ö., Halldórsdóttir, K. L., & Crowe, K. (2024). A pilot study of stuttering treatment for older children (STOC). *Journal of Fluency Disorders*, in press.

[35] Onslow, M., Packman, A., & Harrison, E. (2003). *The Lidcombe Program of early stuttering intervention: A clinician's guide*. Pro-Ed.

[36] Ingham, R. J., Bakker, K., Ingham, J. C., Kilgo, M., & Moglia, R. (2005). *Stuttering Measurement System* (SMS software). http://www.speech.ucsb.edu/

[37] Martin, R. R., Haroldson, S. K., & Triden, K. A. (1984). Stuttering and speech naturalness. *Journal of Speech and Hearing Disorders. 49*(1), 53–58. https://doi.org/10.1044/jshd.4901.53

[38] Einarsdóttir, J. T., Halldórsdóttir K. L., Leosdóttir, B., & Bjarnadóttir, B. (2014). *STUREN homepage-Icelandic samples.* Retrieved from www.sturen.herokuapp.com

[39] Bland, J. M., & Altman, D. G. (1995). Calculating correlation coefficients with repeated observations: Part 1 – correlation within subjects. *BMJ*, 310, 446.

[40] Ratner, B. (2009). The correlation coefficient: Its values range between +1/−1, or do they? *Journal of Targeting, Measurement and Analysis for Marketing, 17*(2), 139–142. https://doi.org/10.1057/jt.2009.5

[41] Benjamini, Y., & Hochberg., Y (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, Series B. 57(1), 289–300. https://doi:10.1111/j.2517-6161.1995.tb02031.x. MR 1325392

[42] Leósdóttir, B. (2014). Heildarmat á upplifun barna sem stama. Þýðing og staðfærsla OASES-S. [Overall assessment of the speaker's experience of stuttering (OASES): translation]. Master thesis. University of Iceland. http://hdl.handle.net/1946/18522

[43] Gresham, F. M., Gansle, K. A., Noell, G. H., Cohen, S., & Rosenblum, S. (1993). Treatment integrity of school-based behavioral intervention studies: 1980-1990. *School Psychology Review*, *22*(2), 254–272.

[44] Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review*, *11*(3), 247–266. https://doi.org/10.1016/0272-7358(91)90103-2

[45] Swift, M. C., Jones, M., O'Brian, S., Onslow, M., Packman, A., & Menzies, R. (2016). Parent verbal contingencies during the Lidcombe Program: Observations and statistical modelling of the treatment process. *Journal of Fluency Disorders, 47,* 13–26. https://doi.org/10.1016/j.jfludis.2015.12.002

[46] Trajkovski, N., Andrews, C., Onslow, M., Packman, A., O'Brian, S., & Menzies, R. (2009). Using syllable-timed speech to treat preschool children who stutter: A multiple

baseline experiment. *Journal of Fluency Disorders, 34*(1), 1–10.

[47] Donaghy, M., O'Brian, S., Onslow, M., Lowe, R., Jones, M., & Menzies, R. G. (2020). Verbal contingencies in the Lidcombe Program: A noninferiority trial. *Journal of Speech, Language, and Hearing Research*, *63*(10), 3419–3431. https://doi.org/10.1044/2020_JSLHR-20-00155

[48] Swift, M. C., O'Brian, S., Hewat, S., Onslow, M., Packman, A., & Menzies, R. (2011). Investigating parent treatment in the Lidcombe Program. *International Journal of Speech-Language Pathology*, *13*(4), 308–316. https://doi.org/10.3109/17549507.2011.550692

[49] Franken, M.-C., & Kielstra-Van der Schalk, C. J., & Boelens, H. (2005). Experimental treatment of early stuttering: A preliminary study. *Journal of Fluency Disorders, 30,* 189–199. https://doi.org/10.1016/j.jfludis.2005.05.002

[50] Greenhaus, J. H., Collins, K. M., & Shaw, J. D. (2003). The relation between work–family balance and quality of life. *Journal of Vocational Behavior*, *63*(3), 510–531.

[51] Goodhue, R., Onslow, M., Quine, S., O'Brian, S., & Hearne, A. (2010). The Lidcombe program of early stuttering intervention: Mothers' experiences. *Journal of Fluency Disorders, 35*(1)*,* 70–84. https://doi.org/10.1016/j.jfludis.2010.02.002

[52] Baldwin, S. A., & Imel, Z. E. (2013). Therapist effects: Findings and methods. Be*rgin and Garfield's handbook of psychotherapy and behavior change, 6,* 258-297.

[53] Harn, B., Parisi, D., & Stoolmiller, M. (2013). Balancing fidelity with flexibility and fit: What do we really know about fidelity of implementation in schools? *Exceptional Children*, 79(3), 181–193. https://doi.org/10.1177/001440291307900204

[54] Borelli, B., Sepinwall, D., Ernst, D., Bellg, A. J., Czajkowski, S., Breger, R., et al. (2005). A new tool to assess treatment fidelity and evaluation of treatment fidelity

across 10 years of health behavior research. *Journal of Consulting and Clinical Psychology,73(5),* 852–860. https://doi.org/10.1037/0022-006X.73.5.852

[55] Peterson, L., Homer, A. I., & Wonderlich, S. A. (1982). The integrity of independent variables in behavior analysis. *Journal of Applied Behavior Analysis, 15*(4), 477–492. https://doi.org/10.1901/jaba.1982.15-477

[56] Perepletchikova, F. (2011). On the topic of treatment integrity. *Clinical Psychology: Science and Practice*, *18*(2), 148–153. https://doi.org/10.1111/j.1468-2850.2011.01246.x

Appendix

Quality

Is the conversation “flowing”?

1. The conversation is flowing, and the child responds with a full-length answer.
2. The conversation is in most scenarios “flowing” with some exceptions.
3. The conversation is mostly forced with some exceptions (the child is often not answering with full length answers)
4. The child is not taking part in the conversation (mostly yes and no answers)

Responsiveness

Is the child happy to take part in the intervention?

1. Very happy – conversation good
2. Mildly happy – conversation fine
3. Not happy to take part in the intervention/conversation.

Table 1
Participant Information

Participant	Age of stuttering onset (years)	Age at start of intervention (years; months)[b]	%SS at baseline	SPM at baseline	Naturalness at baseline[a]	Severity at baseline	Completed intervention
P1	3	10;8	2.0	192.6	4.0	4.0	Yes
P2	10	13;0	6.3	157.7	6.0	6.0	No
P3	2	13;2	8.1	155.4	5.7	5.7	Yes
P4	3	10;6	17.7	82.6	9.0	9.0	Yes
P5	6	9;5	2.5	154.9	3.0	3.0	Yes
P6	3	10;4	2.3	135.5	3.0	4.0	No
P7	3	13;9	17.0	87.2	7.0	7.0	No

Note. %SS = proportion of stuttering; SPM = syllables per minute; [a] Lower scores mean more natural speech; [b] Lower scores mean less severe stuttering.

Table 2
FOI Results for each Participant for Parent-and Child-Managed Intervention Sessions

Participant	Dosage Parent	Dosage Child	Adherence Parent	Adherence Child	Quality Parent	Quality Child	Responsiveness Parent	Responsiveness Child
P1	89.4%	77.3%	75.8%	67.3%	2.1	2.2	1.2	1.2
P2	64.1%	58.5%	44.9%	32.6%	2.1	2.1	1.3	1.4
P3	95.5%	94.7%	75.3%	75.3%	1.2	1.1	1.2	1.1
P4	97.7%	96.5%	80.9%	65.0%	1.1	1.0	1.1	1.0
P5	88.0%	82.0%	57.0%	62.1%	1.7	1.7	1.0	1.1
P6	84.7%	77.0%	60.1%	12.8%	2.0	1.8	1.1	1.1
P7	81.9%	77.4%	58.0%	56.7%	1.7	1.6	1.1	1.1
Mean	85.9%	80.5%	64.6%	53.1%	1.7	1.6	1.1	1.1

Table 3

Correlation Between Dosage and Intervention Outcome Measurements

	N	Mean %SS	Mean SPM	Mean Naturalness	Mean Severity
Total correlation	28	-.30	.31	-.01	-.34
Parent correlation	14	-.27	.22	-.09	-.37
Child correlation	14	-.35	.45	.04	-.35

Note. Each participant has two measures, one for treatment 1 and one for treatment 2. Therefore, there are 14 observations for the 7 subjects for parent on one hand and 14 for child on the other.

Table 4

Correlation Between Adherence and Intervention Outcome Measurements

Parent managed intervention session

Participants	N	%SS	SPM	Naturalness	Severity
P1	83	-.30*	-.05	-.34*	-.43*
P2	88	-.34*	-.13	-.22*	-.28*
P3	69	.11	-.11	-.17	-.17
P4	177	-.24*	-.04	-.38*	-.29*
P5	86	-.40*	-.07	-.16	-.41*
P6	101	-.29*	-.13	-.04	-.27*
P7	160	-.53*	.59*	-.53*	-.53*
All	764	-.45*	.27*	-.17*	-.51*

Child managed intervention sessions

Participants	N	%SS	SPM	Naturalness	Severity
P1	65	-.59*	.09	-.70*	-.67*
P2	79	-.27*	.22	-.06	-.16
P3	75	-.41*	.09	-.40*	-.49*
P4	170	-.25*	-.09	-.40*	-.27*
P5	79	-.43*	-.11	-.15	-.41*
P6	88	-.03	-.05	-.02	-.15
P7	151	-.57*	.47*	-.57*	-.57*
All	707	-.39*	.30*	-.24*	-.48*

Note. *$p < .05$ adjusted with the FDR-controlling procedure [41]

Table 5

Correlation Between Quality and Intervention Outcome Measurements

Parent-managed intervention sessions

Participants	N	%SS	SPM	Naturalness	Severity
P1	83	-.32**	-.24**	-.34**	-.28**
P2	58	-.21*	-.31**	-.02	-.27**
P3	69	-.21*	-.33**	-.12	-.17**
P4	176	-.07	-.16**	-.22**	-.20**
P5	84	-.10	-.02	-.17*	-.13
P6	102	.11	-.42**	.08	.15*
P7	160	.60**	-.76**	.60**	.60**
All	732	.39**	-.54**	-.21**	.25**

Child-managed intervention sessions

Participants	N	%SS	SPM	Naturalness	Severity
P1	65	-.07	-.42**	.07	-.04
P2	55	-.04	-.34**	.00	.03
P3	74	-.07	-.16	-.05	.02
P4	170	-.10	-.14**	-.17**	-.15**
P5	76	-.05	-.41**	-.20*	-.04
P6	87	-.01	-.34**	-.05	.04
P7	150	.49**	-.56**	.49**	.49**
All	677	.48**	-.54**	-.21**	.24**

Note. *p < .05 adjusted with the FDR-controlling procedure [41]

Table 6

Correlation Between Responsiveness and Intervention Outcome Measurements

Parent managed intervention sessions

Participants	N	%SS	SPM	Naturalness	Severity
P1	83	-.04	-.09	-.08	-.10
P2	58	.13	-.23*	.35**	.15
P3	69	-.13	-.45**	-.11	-.14
P4	176	-.11*	-.15**	-.28**	-.25**
P5	84	-.03	-.02	-.07	-.03
P6	102	-.13	.23**	.06	-.06
P7	160	.29**	-.46**	.29**	.29**
All	732	.02	-.46**	-.02	-.03

Child managed intervention sessions

Participants	N	%SS	SPM	Naturalness	Severity
P1	65	-.04	-.15	.07	-.06
P2	55	-.08	-.09	-.12	-.17
P3	74	-.06	-.12	-.06	-.03
P4	170	-.10	-.18**	-.19**	-.15**
P5	76	.06	-.36**	-.11	.05
P6	87	-.14	-.02	-.05	-.09
P7	150	.26**	-.19**	.26**	.26**
All	677	.05	-.22**	-.07	-.03

Note. **p* < .05 adjusted with the FDR-controlling procedure [41]

Table 7
Percentage Difference in Dosage for Children Based on Completion Status

	Participant	Dosage parent	Dosage child
Children who completed STOC	P1	89.4%	77.3%
	P3	95.5%	94.7%
	P4	97.7%	96.5%
	P5	88.0%	82.0%
	Mean	**92.7%**	**87.6%**
Children who did not completed STOC	P2	64.1%	58.5%
	P6	84.7%	77.0%
	P7	81.9%	77.4%
	Mean	**76.9%**	**70.9%**

Table 8

Mean, Standard Deviation, and CI Intervals for Adherence, Quality, and Responsiveness Based on Completion Status

		Parent			**Child**		
		N	Mean (SD)	CI [Lower, Upper]	N	Mean (SD)	CI [Lower, Upper]
Adherence	Completed	412	74.0 (38.0)	± 3.7 [70.1, 77.7]	386	66.8 (41.5)	± 4.1 [62.7, 70.9]
	Did not complete	322	56.2 (35.0)	± 3.8 [52.4, 60.0]	293	41.7 (38.0)	± 4.3 [37.3, 46.0]
Quality	Completed	412	1.46 (0.68)	± 0.07 [1.40, 1.78]	385	1.38 (0.66)	± 0.07 [1.32, 1.45]
	Did not complete	321	1.87 (0.82)	± 0.09 [1.78, 1.96]	292	1.75 (0.76)	± 0.09 [1.66, 1.84]
Responsiveness	Completed	412	1.12 (0.34)	± 0.03 [1.09, 1.15]	385	1.07 (0.29)	± 0.03 [1.04, 1.10]
	Did not complete	321	1.13 (0.38)	± 0.04 [1.09, 1.17]	292	1.12 (0.35)	± 0.04 [1.08, 1.16]

Note. SD = standard deviation; CI = confidence intervals

Appendix A

Ritrýnd grein birt 13. júlí 2020

Orðaforðakennsla með sögulestri fyrir börn með málþroskaröskun

Sigrún Alda Sigfúsdóttir, Jóhanna T. Einarsdóttir,
Þorlákur Karlsson og Íris Ösp Bergþórsdóttir

► Abstract ► Um höfundana ► About the authors ► Heimildir

Tilgangur rannsóknarinnar var að bera saman áhrif beinnar og óbeinnar orðaforðakennslu hjá börnum með málþroskaröskun. Einkenni málþroskaröskunar er slök færni í tungumálinu, bæði í málskilningi og máltjáningu. Beina orðaforðakennslan fólst í að lesa sögubók og skoða hvort börn lærðu ný orð með því að útskýra og vinna sérstaklega með ákveðin markorð sem komu fyrir í textanum. Við óbeina kennslu var sama bók lesin án þess að staldra við markorðin eða útskýra þau. Þátttakendur voru tveir, báðir í elsta árgangi í leikskóla, og höfðu niðurstöður málþroskamælinga fyrir íhlutun sýnt slaka færni, bæði í málskilningi og máltjáningu. Kennslan fór fram í leikskóla barnanna fjórum sinnum í viku, í sex vikur. Niðurstöður leiddu í ljós að góður árangur náðist með þann orðaforða sem kenndur var með beinni kennslu. Orðaforði barnanna jókst hins vegar mun minna við óbeina kennslu. Sú þekking sem börnin höfðu tileinkað sér að lokinni íhlutun hélst að nokkru leyti mánuði eftir að íhlutun lauk. Mikilvægt er að lesa fyrir leikskólabörn og skapa aðstæður þar sem markvisst er verið að kenna ný orð. Jafnframt er nauðsynlegt að huga sérstaklega vel að börnum með slaka málfærni og auðvelda þeim að hlusta á sögu með því að útskýra orð jafnóðum. Þessar niðurstöður gefa vísbendingar um að til að auka orðaforða barna við sögulestur þurfi að útskýra ný orð sérstaklega. Foreldrar, kennarar og talmeinafræðingar geta ekki gert ráð fyrir að börn tileinki sér ný orð með því að heyra þau lesin í sögubók og geti sér til um þýðingu þeirra út frá samhengi.

Efnisorð: Bein og óbein orðaforðakennsla, málþroski, málþroskaröskun, lestur sögubóka

Inngangur

Málfærni leikskólabarna er mjög breytileg, sum börn eru leikin við að orða hugsun sína og þekkja mikinn fjölda orða en önnur þekkja færri orð og eiga erfiðara með að orða það sem þau eru að hugsa. Talið er að 9% nemenda séu með málþroskaröskun (e. Developmental Language Disorder, DLD) (Law, 2019). Málþroskaröskun getur haft víðtæk áhrif bæði á nám og samskipti (Paul, Norbury og Gosse, 2018) en rannsóknir hafa sýnt að málfæri barna við fimm ára aldur spáir fyrir um árangur þeirra í námi síðar meir (Catts, Fey, Tomblin og Zhang, 2002; Jóhanna T. Einarsdóttir, Amalía Björnsdóttir og Ingibjörg Símonardóttir, 2016). Þess vegna er brýnt að veita börnum sem greinast með málþroskaröskun viðeigandi aðstoð eins fljótt og hægt er og efla málþroska þeirra leikskólabarna sem standa höllum fæti. Að lesa sögur fyrir börn hefur verið talin góð leið til að auka orðaforða og efla málfærni þeirra (Paul o.fl., 2018). Markviss orðaforðakennsla með lestri sögubóka (e. storybook reading) hefur talsvert verið rannsökuð erlendis en lítið hér á landi. Aðferðin byggist á að við lestur sögubóka eru valin ákveðin orð (hér eftir kölluð markorð) sem börnin eru líkleg til að eiga erfitt með að skilja og þau orð útskýrð sérstaklega. Í þessari rannsókn verður athugað hvort hægt sé að auka orðaforða barna með málþroskaröskun með því að lesa sögubók og beita markvissri orðaforðakennslu. Orðanám barnanna var kannað, annars vegar með því að lesa sögu og útskýra og vinna sérstaklega með markorðin samkvæmt aðferð beinnar kennslu (e. explicit instruction), og hins vegar án þess að útskýra markorðin sérstaklega, eða með óbeinni kennslu (e. implicit instruction).

Hvað er málþroskaröskun?

Barn sem sýnir marktækt slakari færni í málþroska en jafnaldrar gæti verið með málþroskaröskun. Við greiningu er stuðst við niðurstöður mælinga á formlegum málþroskaprófum en einnig við óformlegar athuganir og mat sérfræðinga (Law, 2019). Málþroskapróf athuga bæði skilning og tjáningu tungumálsins og er þeim skipt niður í þætti sem varða innihald, form og notkun þess. Merkingarfræði (e. semantics) tekur til innihalds tungumálsins, hljóðkerfisfræði (e. phonology), setningafræði (e. syntax) og orðhlutafræði (e. morphemics) snúa að formi þess og að lokum nær málnotkun (e. pragmatics) yfir notkun málsins (Paul o.fl., 2018). Barn með málþroskaröskun á oft í erfiðleikum með alla þætti tungumálsins þótt töluverður einstaklingsmunur og breytileiki sé innan hópsins. Birtingarmynd röskunarinnar lýsir sér í að barnið beitir ekki málfræðireglum rétt, það einfaldar setningar og orðaforðinn er einsleitur. Málnotkun barna með málþroskaröskun birtist í erfiðleikum með að halda sér við umræðuefni, að skiptast á að tala, að segja frá á skipulegan hátt og fylgja almennum reglum sem einstaklingar tileinka sér í samskiptum. Tvítyngt eða fjöltyngt barn getur einnig verið með málþroskaröskun og kemur hún þá fram í öllum tungumálum sem það tileinkar sér.

Málþroskaröskun hefur verið kölluð ýmsum nöfnum, eins og seinn málþroski (e. language delay), málhömlun (e. language impairment) eða sértæk málþroskaröskun (e. specific language disorder). Árið 2016 vann alþjóðlegur hópur, svokallaður CATALISE-hópur, að því að sammælast um viðmið fyrir frávik í málþroska. Þátttakendur voru 57 sérfræðingar úr tíu ólíkum starfsstéttum, svo sem talmeinafræðingar, sálfræðingar, kennarar, læknar og einnig fulltrúar úr fjölskyldum barna sem voru með frávik í málþroska. Hópurinn ákvað að nota hugtakið málþroskaröskun (e. DLD) ef frávik í málþroska væru vegna óþekktra orsaka. Hópurinn lagði jafnframt til að hugtakið málröskun (e. language disorder) yrði notað yfir frávik í máli sem fylgdu þekktri fötlun og þá væri fjallað um málröskun tengda einhverfu eða málröskun tengda Downs-heilkenni (Bishop, Snowling, Thompson og Greenhalgh, 2016, 2017). Ákvarðanir hópsins hafa haft víðtæk áhrif á alþjóðlega umfjöllun vísindamanna og greiningu á málþroskafrávikum barna og unglinga. Greiningarviðmið voru einfölduð þannig að ef barn greinist á eftir í máli þrátt fyrir að hafa fengið viðhlítandi málörvun úr umhverfinu er talið að það sé með málþroskaröskun. Ekki þurfa að liggja fyrir þroskamælingar með greindarprófum eins og áður var skilyrði þegar um var að ræða sértæka málþroskaröskun. Alþjóðlegir hópar hafa jafnframt unnið að því að auka þekkingu og vitund um málþroskaröskun, meðal annars undir merkjum RADLD – Raising Awareness of Developmental Language Disorders (RADLD, e.d.).

Orðaforði barna með málþroskaröskun

Eitt megineinkenni málþroskaröskunar er slakur orðaforði. Orðaforði (e. vocabulary) er skilgreindur sem sá fjöldi orða sem einstaklingur hefur á valdi sínu (Hiebert og Kamil, 2005). Bæði er átt við hversu mörg orð einstaklingurinn þekkir, þ.e. stærð orðaforðans, og hversu vel hann skilur hvert orð, eða dýpt orðaforðans (McGregor, Oleson, Bahnsen og Duff, 2013).

Börn með málþroskaröskun eru oft sein til að segja fyrsta orðið og þau eru lengur að læra ný orð en jafnaldrar þeirra með dæmigerðan málþroska (McGregor o.fl., 2013). Þau einskorða sig við algeng orð í tali og nota sjaldan fátíð orð þrátt fyrir að þekkja þau (Nation og Snowling, 2004). Gera má ráð fyrir að þau þurfi að heyra nýtt orð a.m.k. tvisvar sinnum oftar en börn með dæmigerðan málþroska áður en þau skilja orðið og nota upp á eigin spýtur (Paul o.fl., 2018). McGregor o.fl. (2013) fylgdust með breytingum á orðaforða 502 barna með málþroskaröskun.

Þau voru prófuð í öðrum, fjórða, áttunda og tíunda bekk og athugað hversu mörg orð þau gátu nefnt (stærð orðaforðans) og hversu vel þau skilgreindu hvert orð (dýpt orðaforðans). Niðurstöður bentu til þess að börn á skólaaldri með málþroskaröskun hefðu almennt minni orðaforða en jafnaldrar þeirra með dæmigerðan málþroska. Orðaforði þátttakenda reyndist takmarkaður, bæði hvað varðar stærð og dýpt og hélst munurinn alla skólagönguna.

Þá hefur einnig verið sýnt fram á að börn með málþroskaröskun eiga í erfiðleikum með margræðni orða og myndmál. Þau njóta þess síður að láta lesa fyrir sig og með því missa þau af tækifæri til að læra ný orð og tileinka sér málfræði og setningagerð sem kemur fram í sögubókum (Paul o.fl., 2018).

Þegar talað er um orðaforða barna er óhjákvæmilegt að minnast á lestur (e. reading) en bein tengsl eru á milli orðaforða og lesskilnings (Paul o.fl., 2018). Börn þurfa að skilja merkingu orða og setninga sem þau lesa til að ná samhengi og skilja textann. Laufer og Ravenhorst-Kalovski (2010) sýndu að nemendur í framhaldsskóla þyrftu að þekkja 98% orða í texta til að geta skilið hann og tileinkað sér án utanaðkomandi aðstoðar. Fari þessi tala niður í 95% þurfa flestir nemendur aðstoð, til dæmis orðabók, til að skilja textann.

Að kenna eða læra ný orð

Ýmsar leiðir eru árangursríkar til að auka orðaforða barna. Meðalbarn lærir merkingu mörg hundruð orða á ári, þótt það geti verið gríðarlegur einstaklingsmunur á hvað börn þekkja mörg orð. Ný orð lærast oft í samræðum og í samskiptum við aðra (Rice, Buhr og Nemeth, 1990). Börn með slakan orðaforða við upphaf skólagöngu þurfa að bæta við sig nokkur hundruð orðum aukalega á ári til að eiga möguleika á að ná jafnöldrum sínum (Biemiller og Boote, 2006).

Í safngreiningu (e. meta-analysis) Marulis og Neuman (2010) skoðuðu þeir 67 rannsóknir á orðaforðakennslu ungra barna. Í þessum rannsóknum var algengast að nota sögulestur eða hliðstæða aðferð sem nefnd er samræðulestur (e. dialogic reading) til að auka orðaforða barna og orðanám barnanna borið saman með beinni kennslu annars vegar og óbeinni kennslu hins vegar. Þeirra niðurstaða var að bein kennsla, með því að útskýra markorð sérstaklega, væri árangursrík leið til að auka orðaforða barna. Þessi aðferð var talin árangursríkari en óbein kennsla þar sem reiknað var með að börn lærðu merkingu orða út frá samhengi. Einnig gafst vel að blanda saman beinni og óbeinni kennslu þannig að markorðin voru fyrst lesin í samhengi og síðan útskýrð beint, rifjuð upp og unnið með þau sérstaklega í verkefnum. Aðrar rannsóknir hafa einnig sýnt jákvæð áhrif beinnar kennslu (Beach, Sanchez, Flynn og O'Connor, 2015; Biemiller og Boote, 2006; McKeown og Beck, 2014). Hugmyndafræði beinnar kennslu kom fyrst fram á sjónarsviðið í byrjun níunda áratugar 20. aldar. Fyrst var aðferðin sérstaklega hugsuð fyrir börn á grunnskólaaldri (Beck, Perfetti og McKeown, 1982; McKeown, Beck, Omanson og Perfetti, 1983; McKeown, Beck, Omanson og Pople, 1985) en í byrjun 21. aldar var aðferðin aðlöguð að leikskólabörnum (Beck og McKeown, 2007).

Í safngreiningu Marulis og Neuman (2010) var tímalengd, tímafjöldi og lengd hvers íhlutunartíma einnig athuguð. Ekki fékkst einhlít niðurstaða um hvað væri árangursríkast enda fjölmargar breytur sem komu við sögu. Áhrifastærð (e. effect size) voru til dæmis meiri ef þjálfunin tók styttri tíma í vikum talið og ef tímar voru fáir. Þegar skoðuð var lengd hvers þjálfunartíma kom í ljós að lengri tímar voru ekki árangursríkari en styttri. Þjálfunartími sem var styttri en 20 mínútur, að lágmarki 7 mínútur, gaf ekki marktækt verri niðurstöðu en lengri tími.

Riches, Tomasello og Conti-Ramsden (2005) athuguðu hversu fljótt og vel börn með sértæka málþroskaröskun tileinkuðu sér nýjar sagnir og báru saman við samanburðarhóp með eðlilegan málþroska. Þau komust að því að börn með sértæka málþroskaröskun þurftu bæði að heyra nýja orðið oftar og yfir lengri tíma en börn í samanburðarhópnum. Enn fremur var börnum með sértæka málþroskaröskun hættara við að gleyma orðunum eftir að íhlutun lauk.

Í skýrslu National Reading Panel (National Institute of Child Health and Human Development, 2000) var einkum þrennt sem talið var árangursríkt við að auka orðaforða barna; í fyrsta lagi að nota sjónrænan stuðning, í öðru lagi að heyra orð endurtekið og í þriðja lagi að kenna orð í fjölbreyttu samhengi. Sögulestur gefur möguleika á að nýta alla þessa þætti. Auðvelt er að endurtaka með því að lesa sömu bók aftur og aftur og ræða um ákveðin orð, sögubækur fyrir yngri börn eru yfirleitt mikið myndskreyttar sem getur nýst sem sjónrænn stuðningur og orð koma fyrir í fjölbreyttu samhengi sem getur dýpkað skilning á merkingu þeirra.

Orðaforðakennsla með sögulestri

Niðurstöður rannsókna sýna að hægt sé að auka orðaforða barna með því að lesa fyrir þau sögur (Walsh og Blewitt, 2006; Wasik og Bond, 2001). Rannsóknir á málfærni íslenskra leikskólabarna hafa sýnt að börn sem oft er lesið fyrir búa yfir ríkulegri orðaforða en þau sem sjaldan er lesið fyrir (Hrafnhildur Ragnarsdóttir, 2015) og börn sem lesið er fyrir daglega hafa betri mál- og hljóðvitund en börn sem lesið er sjaldnar fyrir (Amalía Björnsdóttir, Ingibjörg Símonardóttir og Jóhanna Einarsdóttir, 2003).

Áhrif sögulestrar á orðanám barna hafa verið skoðuð í mismunandi útfærslum í mörgum rannsóknum (Beck, McKeown og Kucan, 2013; Biemeller og Boote, 2006; Coyne, McCoach og Kapp, 2007; Wilkinson og Houston-Price, 2013). Aðferðin Orðaspjall hefur verið þróuð fyrir íslenska leikskóla (Árdís Hrönn Jónsdóttir, 2013) og byggist á aðferð Beck o.fl. (2013) um markvissa kennslu orða með sögulestri. Markmiðið er að efla skilning barna á samfelldum texta og auka orðaforða þeirra með markvissum hætti. Kennari les sögubók með börnunum og útskýrir merkingu orða sem koma fyrir í bókinni. Eftir lesturinn fer kennarinn yfir söguþráðinn og hvetur börnin til að velta sögunni fyrir sér með samræðum um bókina. Við val á orðum er stuðst við flokkun Beck o.fl. (2013) á þremur mismunandi lögum orðaforðans (e. three tier model of vocabulary words). Algeng orð úr talmáli sem þurfa sjaldnast sérstaka kennslu eru í fyrsta lagi orðaforðans (e. tier one). Þetta eru orð sem flokkast undir grunnorðaforða líkt og *sofa, maður, lítill, ég*. Algeng orð í ritmáli og námsbókum eru í millilagi orðaforðans (e. tier two). Þetta eru oft sjaldgæfari orð sem eru notuð yfir algeng hugtök, líkt og *lúinn* í stað *þreyttur* og *kjökra* í stað *gráta*. Í þriðja laginu (e. tier three) eru sjaldgæf og sértæk orð sem koma einkum fyrir á tilteknum sérsviðum, t.d. í vísindum, tæknifræði og atvinnugreinum. Má þar nefna orð líkt og móberg, *málþroskafrávik* og *meðaltal*. Í Orðaspjalli eru valin orð úr millilagi orðaforðans til að útskýra sérstaklega.

Biemiller og Boote (2006) báru saman aðferð beinnar kennslu og óbeinnar við að auka orðaforða grunnskólabarna með sögulestri. Þar var bók lesin endurtekið og annars vegar kannað hversu vel börnin tileinkuðu sér markorð sem voru útskýrð beint og hins vegar án beinnar útskýringar. Rannsóknin stóð yfir í eina viku og voru valin 24 markorð fyrir hvern aldurshóp. Að meðaltali náðu börnin merkingu 12% markorðanna með því að heyra orðin endurtekið en þeim fjölgaði upp í 22% við að fá beina útskýringu á þeim. Rannsóknin var síðan endurtekin ári seinna með fleiri þátttakendum og þá voru öll orð útskýrð sem talið var að börnin skildu ekki. Seinni rannsóknin sýndi betri árangur en sú fyrri. Börnin lærðu þá að meðaltali 40% af markorðunum. Coyne og félagar (2007) notuðu fyrrnefnda aðferð hjá börnum á leikskólaaldri (5–6 ára). Þeir báru saman orðanám barna í skóla þar sem stór hluti barna var í áhættu varðandi lestrarerfiðleika. Þátttakendum var skipt af tilviljun í tvo hópa og voru 3–4 í hvorum hópi. Rannsóknin tók eina viku. Hvor hópur heyrði söguna um grísina þrjá lesna þrisvar sinnum. Valin voru sex markorð (tvö nafnorð, tvö sagnorð og tvö lýsingarorð) úr sögunni sem þóttu mikilvæg og ólíklegt að börnin þekktu. Þrjú orð voru kennd með beinni útskýringu með því að kynna þau sem töfraorð áður en lesturinn hófst. Þegar eitthvert markorðanna kom síðan fyrir í setningu var hún lesin aftur og orðið útskýrt, hvort sem börnin tóku eftir því eða ekki. Eftir söguna var flett í gegnum bókina og markorðin rifjuð upp, þau endurtekin og farið yfir hvernig þau komu fyrir í sögunni. Þrjú orð voru ekki kennd sérstaklega og þá var sagan lesin án þess að stöðva við markorðin eða kynna þau til leiks. Niðurstöður rannsóknarinnar sýndu að börnin lærðu marktækt fleiri orð með beinni kennslu en með óbeinni kennslu og var sú kunnátta enn til staðar átta vikum eftir að íhlutun lauk.

Tvær rannsóknir hafa verið gerðar á Íslandi þar sem könnuð voru áhrif beinnar kennslu með sögulestri á orðskilning fimm ára tvítyngdra barna. Báðar sýndu að börnin lærðu mun fleiri markorð sem voru útskýrð beint en samanburðarorð sem ekki voru útskýrð (Eyrún Rakel Agnarsdóttir, 2018; Helga Hilmarsdóttir, 2016). Helga Hilmarsdóttir (2016) sýndi enn fremur að þátttakandinn í hennar rannsókn viðhélt þekkingu sinni á orðunum sem kennd voru með beinni kennslu mánuði eftir að rannsókninni lauk. Aðferðin hefur hins vegar ekki áður verið skoðuð með börnum sem þurfa á sérstakri örvun að halda og eru á eftir í málþroska.

Rannsóknarspurningar

Markmið rannsóknarinnar var að skoða aðferð til að auka orðaforða hjá tveimur börnum með málþroskaröskun í elsta árgangi í leikskóla og meta árangurinn af þeirri kennslu. Hún fól í sér að útskýra fyrirfram ákveðin markorð þegar þau komu fyrir í sögubók og vinna með þau að lestri loknum. Leitast var við að svara eftirfarandi spurningum:

- Skilar bein útskýring á markorðum, ásamt kennslu að lestri loknum, betri árangri til að auka orðaforða barnanna en að fá enga útskýringu á markorðum?
- Ef orðaforðinn eykst, helst sú kunnátta að einhverju marki mánuði eftir að kennslu lýkur?
- Hefur orðaforðakennsla með lestri sögubóka áhrif á almenna máltjáningu barnanna eins og hún er metin með málsýnum?

Aðferð

Þátttakendur

Tveir þátttakendur voru í rannsókninni og uppfylltu þeir eftirtalin skilyrði; 1) fæddir og aldir upp á Íslandi og á síðasta ári í leikskóla, 2) sýna slaka færni á málþroskaprófinu TOLD-2P (1,5 staðalfrávik undir meðalgetu jafnaldra) og 3) sýna slaka færni á hljóð- og málvitundarprófinu HLJÓM-2. Valdir þátttakendur voru drengur og stúlka sem verða hér eftir kölluð Tómas og Eva en það eru ekki þeirra réttu nöfn. Börnin voru í sama leikskóla á höfuðborgarsvæðinu.

Tómas fæddist á Íslandi. Hann ólst upp við tvö tungumál, faðir hans talaði við hann íslensku en móðir hans talaði við hann spænsku sem var hennar fyrsta mál. Tómas virtist skilja einhverja spænsku en talaði hana ekki og svaraði á íslensku þegar móðir hans talaði við hann á spænsku. Íslenskan var hans sterkara tungumál. Fyrst fór að bera á málþroskavanda Tómasar um þriggja ára aldur en athuganir með málþroskaprófum sýndu þá slaka færni eða málþroskatölu sem var tæpum 2 staðalfrávikum undir meðalgetu jafnaldra. Niðurstöður á HLJÓM-2 við fimm ára aldur sýndu mjög slaka færni og einnig niðurstöður TOLD-2P. Starfsfólk leikskólans taldi að Tómas ætti í erfiðleikum með að fylgjast með í samverustundum og tók rannsakandi (fyrsti höfundur greinarinnar) sjálfur eftir því þegar hann fylgdist með Tómasi í slíkri stund þar sem saga var lesin. Tómas truflaði ekki mikið heldur virtist hann vera annars hugar og ekki fylgjast með því sem fram fór. Tómas fellur samkvæmt skilgreiningu undir að vera með málþroskaröskun þar sem hann sýnir veruleg frávik í báðum þeim tungumálum sem eru til staðar í umhverfi hans. Tómas var 6 ára þegar íhlutunin átti sér stað.

Eva er einnig fædd á Íslandi og á íslenska foreldra. Niðurstöður á málþroskaprófinu TOLD-2P þegar Eva var um fimm ára gömul sýndu færni sem var 1,5 staðalfrávikum undir meðalgetu jafnaldra. Ennfremur sýndu niðurstöður athugunar með HLJÓM-2 slaka færni. Starfsfólk leikskólans taldi að Eva ætti í erfiðleikum með tungumálið. Þegar rannsakandi fylgdist með Evu í sögustund virtist hún eiga erfitt með að sitja kyrr og hlusta á söguna. Yfirleitt fór hún að gera eitthvað annað, fiktaði í hlutum sem hún hafði meðferðis eða reyndi að trufla sessunaut sinn. Eva var rúmlega 5 ára og 6 mánaða þegar íhlutunin fór fram.

Mælingar

Mælingar voru gerðar með sérhönnuðu orðaforðaprófi og með málsýnum. Auk þess var meðferðarheldni (e. fidelity) mæld sérstaklega. Með meðferðarheldni er átt við hvort rannsóknin hafi verið framkvæmd eins og ákveðið var í upphafi, sjá nánar síðar.

Í *Töflu 1* sést yfirlit yfir þær mælingar sem gerðar voru.

Tafla 1. Yfirlit yfir mælingar gerðar meðan á rannsókninni stóð.

Tími mælinga	Mælitæki
Mælingar á grunnskeiði	Sérhannað mælitæki, alls 36 orð Málsýni
Daglegar mælingar eftir hvern kennslutíma	Orð dagsins (markorð eða samanburðarorð), alls 6 orð mæld með sérhönnuðu mælitæki
Vikulegar mælingar	Málsýni
Mælingar um leið og kennslu lauk	Sérhannað mælitæki, alls 36 orð Málsýni
Mælingar mánuði eftir að kennslu lauk	Sérhannað mælitæki, alls 36 orð Málsýni

Sérhannað mælitæki

Útbúið var sérhannað mælitæki til að prófa úr markorðaforðanum sem var alls 36 orð, 18 markorð sem voru útskýrð sérstaklega og 18 orð sem ekki voru útskýrð. Byggt var á rannsókn Coyne og félaga (2007) við hönnun mælitækisins. Mælingar voru fyrir íhlutun, strax eftir íhlutun og að lokum mánuði síðar þegar viðhaldsmælingar voru gerðar, sjá *Töflu 1.* Daglega var spurt um markorðin sex sem tengdust þeirri bók sem lesin var hverju sinni.

Prófun úr markorðaforðanum fór þannig fram að rannsakandi spurði barnið *Hvað þýðir X*? eða *Hvað er X*? Dæmi: *Hvað þýðir kroppur*? eða *Hvað þýðir að vera sorgmæddur*? Tvö stig fengust fyrir rétt svar eins og rannsakandi hafði fyrirfram skilgreint það. Eitt stig fékkst ef barnið sagði einhvern hluta orðsins rétt eða benti. Sem dæmi orðið kroppur, en þá fékk barnið eitt stig ef það benti á einhvern líkamshluta eða nefndi a.m.k. tvo líkamshluta. Þar sem sérhannaða mælitækið var gert sérstaklega fyrir þessa rannsókn var ekki um stöðluð svör að ræða og var það mat rannsakanda hvað teldist rétt og hvað rangt. Öll svör barnsins voru skráð og farið var yfir fyrirgjöf. Fyrir markorðin 36 var mest hægt að fá 72 stig. Til að gera prófunina meira aðlaðandi fyrir börnin skrifaði höfundur orðin á miða, setti þá í skál og börnin drógu af tilviljun og útskýrðu.

Málsýni

Málsýni (e. language sample) eru óformlegar athuganir á málþroska barna og stutt sýnishorn af máltjáningu barns í samtali við aðra. Þau eiga að sýna dæmigerða máltjáningu barnsins við eðlilegar aðstæður og eru mjög oft notuð samhliða stöðluðum prófum. Málsýni eru oft notuð til að mæla árangur af íhlutun þar sem hægt er að taka þau endurtekið. Við úrvinnslu málsýna eru notaðar eftirtaldar mælieiningar: Meðallengd segða (MLS), fjöldi mismunandi orða (FMO), heildarfjöldi orða (HFO) og hlutfall villna. Tali barnanna var skipt í segðir (e. utterance) eftir ákveðnum reglum og er ein segð oftast sambærileg einni setningu Í þessari rannsókn verður stuðst við mælieiningarnar FMO og hlutfall villna. FMO gefur mynd af fjölbreytileika í orðanotkun barna í sjálfsprottnu tali og við hlutfall villna er skoðað hversu mörg orð eru málfræðilega rangt mynduð. (Jóhanna Thelma Einarsdóttir og Þóra Sæunn Úlfsdóttir, 2018). Viðmiðunartölur eru til fyrir þessar mælingar og miða þær við 50 segðir barns. Viðmiðunartölur voru fengnar frá 39 börnum í aldurshópnum 5;6 til 5;11 og frá 25 börnum í aldurshópnum 6;0 til 6;5 ára. Stuðst er við meðaltal en þess ber að geta að staðalfrávik innan hvers aldurshóps fyrir FMO er hátt og ekki hefur verið gefið upp staðalfrávik fyrir hlutfall villna (Jóhanna T. Einarsdóttir og Álfhildur Þorsteinsdóttir, 2015).

Samtals voru 13 málsýni tekin frá hvoru barni sem skiptust þannig að þrjú voru tekin áður en kennsla hófst, sex voru tekin í lok hverrar viku, þrjú að kennslu lokinni og að lokum eitt þegar viðhaldsmælingar voru gerðar mánuði eftir að kennslu lauk. Málsýnin voru tekin við leik og í samtali rannsakanda og þátttakenda. Þau voru tekin upp (hljóð og mynd) með iPhone 7 síma og afrituð samkvæmt handbók um málsýni (Jóhanna Thelma Einarsdóttir og Þóra Sæunn Úlfsdóttir, 2018). Hugbúnaðurinn *Málgreinir* (Jóhanna Thelma Einarsdóttir og Stefán Carl Peiser, 2016) var notaður til að reikna út fjölda mismunandi orða (FMO) og hlutfall villna.

Efni

Við kennsluna voru notaðar sögubækur sem voru valdar þannig að þær væru ekki of langar þar sem aðeins var gert ráð fyrir að hámarki 30–40 mínútum í hvern kennslutíma. Þá þótti einnig mikilvægt að gæta þess að bækurnar væru ekki of erfiðar og að líklegt væri að börnunum þættu þær skemmtilegar og myndu vekja áhuga þeirra. Sex bækur voru notaðar í kennslunni.

Val á orðum

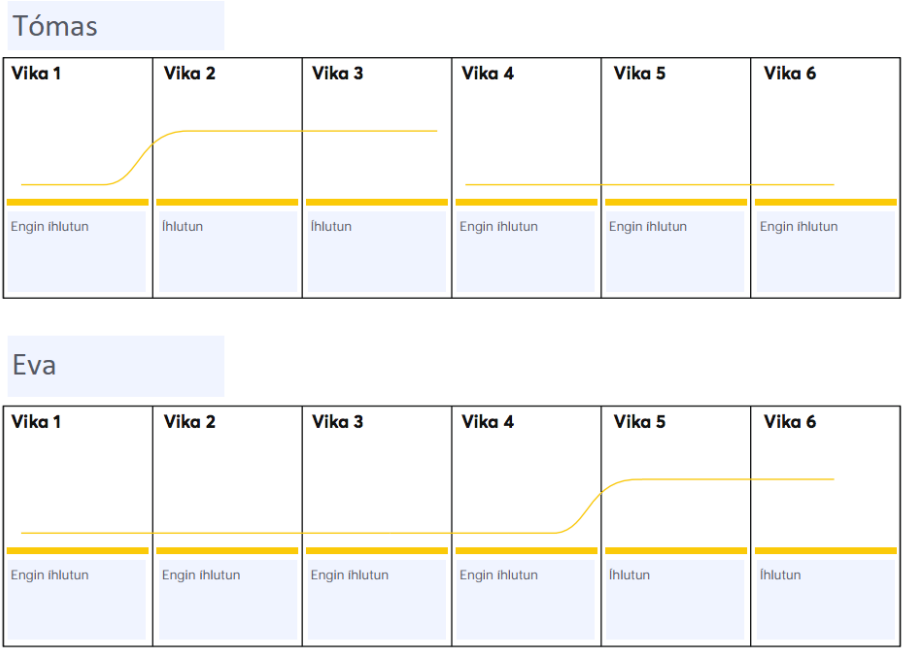
Valin voru sex markorð úr hverri bók, tvö nafnorð, tvö sagnorð og tvö lýsingarorð. Samtals voru markorðin því 36 úr bókunum sex. Af þeim fengu börnin beina útskýringu á 18 orðum en hin 18 voru ekki útskýrð og notuð til samanburðar. Við val á orðum var farið í gegnum hverja bók og fundin sex orð út frá viðmiðun Beck og félaga (2013) um millilag orðaforðans. Í beinni kennslu voru markorðin kennd um leið og þau komu fyrir í textanum og stutt verkefni gerð í lokin en í óbeinni kennslu voru samanburðarorð ekki kennd. Markmiðið með samanburðarorðunum var að kanna hvort þekking barnanna á orðunum ykist með því að heyra þau í samhengi eða hvort beina orðaforðakennslu þyrfti til og að heyra orðin endurtekið. Markorðin voru skilgreind með aðstoð orðabókar og skilgreiningarnar aðlagaðar að aldri barnanna. Útskýring til dæmis á markorðinu ástúð var að þykja vænt um einhvern. Dæmi: Þér þykir vænt um mömmu þína. Í Töflu 2 má sjá upplýsingar um markorð og samanburðarorð úr hverri bók.

Tafla 2. Bækur, markorð/samanburðarorð og kennsluaðferð.

Bók	Markorð og samanburðarorð	Bein eða óbein kennsla
Konungsborna bólubaslið	*Kollur, dónaskapur, kjökra, kemba, fjarlægð, sorgmæddur.*	Bein kennsla
Búkolla	*Ástúð, bál, mæla, míga, ógnarstór, stórstíg.*	Bein kennsla
Letipúkar	*Kroppur, hrúga, nenna, spjalla, varlega, snöggur.*	Bein kennsla
Stór og svolítið pirrandi fíll	*Líkamsrækt, blundur, troðast, mölbrjóta, harkalega, risavaxinn.*	Óbein kennsla
Fjörugt ímyndunarafl	*Félagsskapur, öngþveiti, háma, kremja, glorhungraður, undrandi.*	Óbein kennsla
Gleym-mér-ei	*Hjörð, sólsetur, skrækja, mása, dýrmætt, hugrakkur.*	Óbein kennsla

Rannsóknarsnið

Við greiningu gagna var notað snið margþætts grunnskeiðs (e. multiple-baseline design) sem er ein tegund af einliðasniði (e. single-subject design). Það er notað til að skoða sams konar hegðun hjá nokkrum einstaklingum með því að beita íhlutun á ólíkum tímum (Guðrún Árnadóttir og Þorlákur Karlsson, 2003; Kazdin, 2011). Fyrstu vikuna fengu hvorki Tómas né Eva kennslu, þ.e. bækurnar voru lesnar án þess að staldra nokkuð við ákveðin orð. Í annarri og þriðju viku fékk Tómas kennslu en ekki Eva. Í fjórðu viku fengu hvorki Tómas né Eva kennslu. Í fimmtu og sjöttu viku fékk Eva kennslu en Tómas ekki. Á *Mynd 1* má sjá yfirlit um snið margþætts grunnskeiðs eins og því var beitt í rannsókninni.



Mynd 1. Snið margþætts grunnskeiðs í rannsókninni.

Framkvæmd

Íhlutunaraðferðin byggist á aðferð Coyne og félaga (2007). Rannsókninni var breytt þannig að þátttakendur voru færri og með skilgreinda málþroskaröskun, rannsóknartíminn var lengdur, fleiri bækur voru notaðar og fleiri markorð. Íhlutunin fór fram fjórum sinnum í viku yfir sex vikna tímabil í litlu sérkennsluherbergi á leikskóla barnanna og var rætt við eitt barn í senn. Við beina kennslu tók hver tími um 30–40 mínútur en við óbeina kennslu var hver tími aðeins um 15 mínútur.

Í hverri viku voru lesnar þrjár bækur en þar sem íhlutunin var fjórum sinnum í viku var sama bókin lesin tvisvar, til dæmis á mánudegi og svo aftur á fimmtudegi. Líkt og í rannsókn Coyne og félaga (2007) voru markorðin kynnt fyrir barninu áður en lesturinn hófst og talað um að þetta væru töfraorð. Barnið var beðið um að veita þeim eftirtekt þegar þau kæmu fyrir í sögunni. Ef barnið tók eftir orðinu var setningin lesin aftur og útskýring gefin á markorðinu. Ef barnið tók ekki eftir orðinu var útskýringin samt sem áður gefin og setningin lesin aftur. Dæmi: Ef barnið tók eftir því að orðið *sorgmædd* kom fyrir í sögunni var sú setning lesin aftur og orðið útskýrt á eftirfarandi hátt: *Sorgmæddur þýðir að vera mjög leiður eða dapur.* Farið var yfir öll markorðin á sama hátt og sagan lesin til enda. Eftir hverja sögu var flett í gegnum bókina, markorðin rifjuð upp og unnið lítið verkefni sem fólst í að ræða um markorðin, útskýra þau aftur og barnið teiknaði það sem gerðist í sögunni þegar hvert orð kom fyrir. Að lokum var rætt stuttlega í hvers konar aðstæðum merking orðsins gæti komið fyrir. Þannig var hvert markorð útskýrt aftur. Við óbeina kennslu voru bækurnar lesnar án þess að staldra við markorðin og engin verkefni voru í lok tímans. Þar var verið að kanna hvort börnin áttuðu sig á merkingu orðanna með því að heyra þau í samhengi við textann í sögunni. Þá heyrðu börnin ekki markorðin endurtekin eins og við beina kennslu.

Meðferðarheldni

Meðferðarheldni var könnuð með fjórum mælingum. Þær voru: Magn íhlutunar (e. dosage), fastheldni (e. adherence), gæði meðferðar (e. quality) og svörun (e. responsiveness). Hver þessara mælinga er mikilvæg til að útskýra og sýna á trúverðugan hátt hvað skilar árangri í rannsókninni (Dusenbury, Brannigan, Falco og Hansen, 2003).

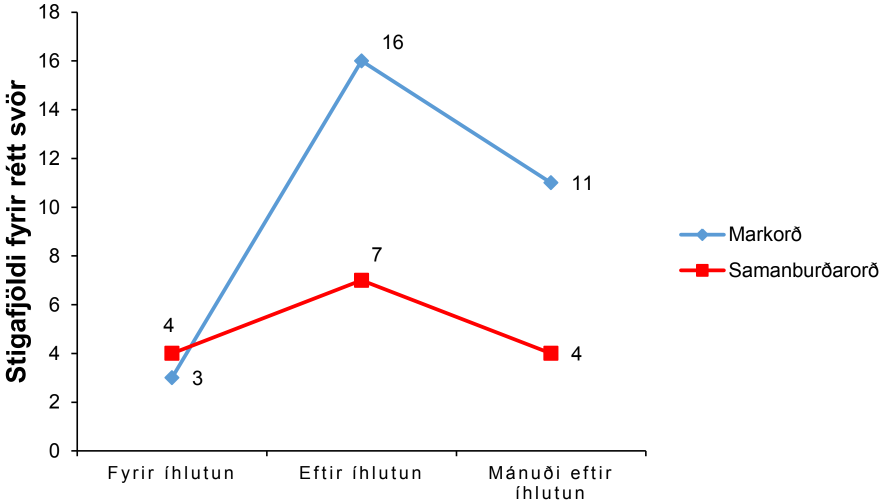
Magn íhlutunar felur í sér mælingu á fjölda raunverulegra skipta sem íhlutun er veitt, lengd hvers skiptis og fjölda daga/vikna sem íhlutun er veitt. Fastheldni er skilgreind sem mælikvarði á hvort íhlutunin var veitt af nákvæmni. Fastheldni var könnuð hér með því að mæla á þriggja punkta kvarða hvort rannsakandi útskýrði markorðin eins og lagt var upp með í beinni kennslu. Einnig var kannað hvort rannsakandi sleppti útskýringu á samanburðarorðunum þegar óbein kennsla fór fram. Kvarðinn var þannig að 1 stóð fyrir *ekki viðunandi*, 2 stóð fyrir *gerir stundum* og 3 stóð *fyrir viðunandi*. Gæði íhlutunar segja til um hversu vel henni var komið til skila eða hversu vel rannsakandinn kom íhlutuninni frá sér og hvernig samskipti rannsakanda og þátttakenda voru. Gæði rannsóknarinnar voru könnuð með því að horfa til þess hvort rannsakandi læsi bókina og útskýrði markorðin þannig að hann næði athygli barnanna og sýndi að hann væri að leggja sig fram. Niðurstöðurnar voru einnig skráðar á sama þriggja punkta kvarða og í mælingunni á fastheldni. Að lokum var svörun við íhlutun athuguð en hún vísar til þess hversu vel þátttakandi brást við íhlutuninni og einnig var skoðað að hve miklu leyti þátttakendur tóku þátt og hvort þeir voru virkir þátttakendur í kennslunni. Svörun var mæld í rannsókninni með sama þriggja punkta kvarða og metið hvort börnin höfðu ánægju af lestrinum eða ekki. Þessi mæling fól í sér að skoða þátttöku barnanna, út frá því hvort þau tóku virkan þátt í umræðunni sem átti sér stað í íhlutuninni. Í þessari mælingu var stigagjöfin þannig að 1 stóð fyrir *hlustar ekki*, *hefur ekki ánægju af lestrinum*, 2 stóð fyrir *hlustar með hálfum huga* og 3 stóð fyrir *hefur ánægju af lestrinum og hlustar með athygli*. Eftir að rannsókninni lauk horfði rannsakandi á öll myndböndin til að meta meðferðarheldni út frá mælingunum fjórum og óháður matsaðili var fenginn til að fara yfir 30% myndbandanna til að meta hið sama. Að lokum voru niðurstöður bornar saman og samræmi skoðað á milli rannsakandans og óháða matsaðilans.

Niðurstöður

Sérhannað mælitæki – markorðaforði

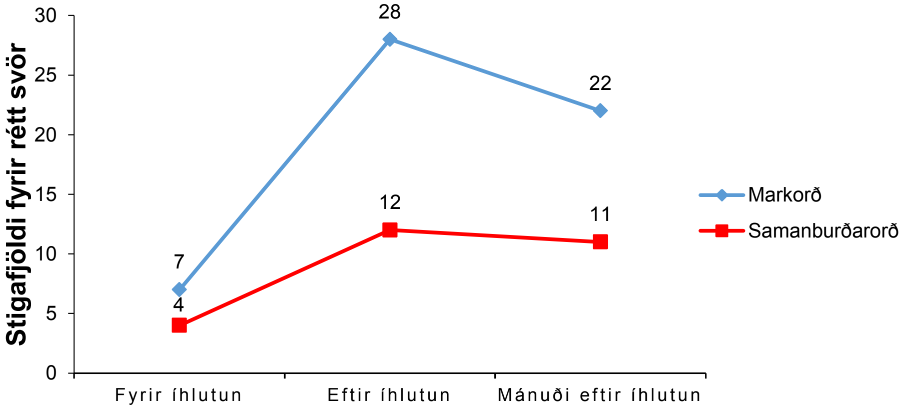
Próf úr öllum markorðaforðanum var lagt fyrir áður en kennslan hófst, strax eftir að henni lauk og um mánuði seinna. Fyrst verða niðurstöður skoðaðar hjá Tómasi og síðan hjá Evu.

Á *Mynd 2* má sjá fjölda mark- og samanburðarorða sem Tómas kunni fyrir íhlutun, strax að henni lokinni og loks mánuði eftir að íhlutun lauk. Eins og sést á myndinni gat Tómas útskýrt mun fleiri markorð eftir íhlutun, bæði strax á eftir og mánuði síðar en fyrir hana, en það átti ekki við um samanburðarorðin.



Mynd 2. Fjöldi réttra svara hjá Tómasi á markorðaforðaprófi fyrir íhlutun, eftir íhlutun og mánuði eftir íhlutun.

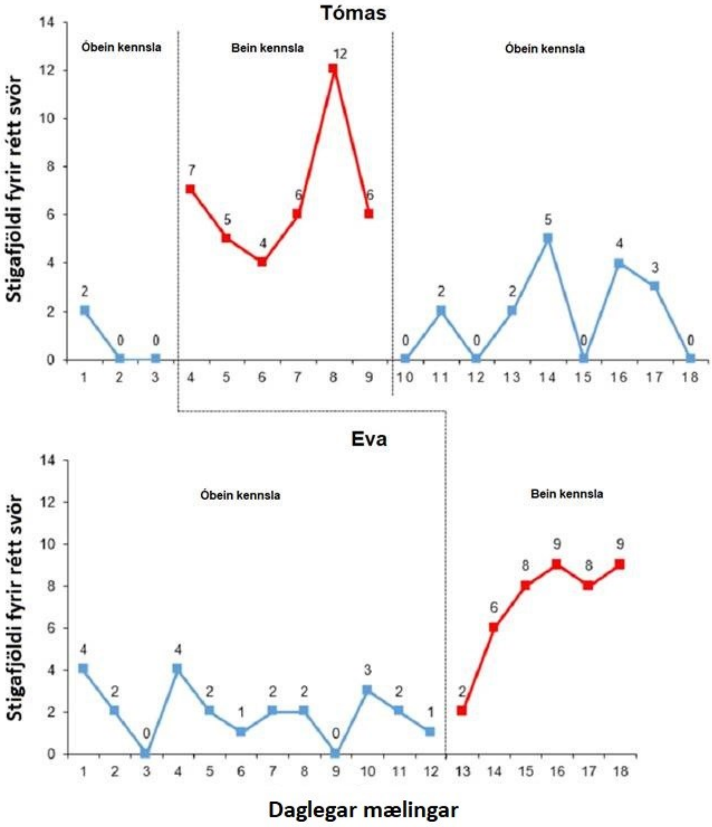
Á *Mynd 3* má sjá fjölda mark- og samanburðarorða sem Eva kunni. Eins og sjá má á myndinni jókst kunnátta Evu strax eftir kennslu, bæði á markorðunum og samanburðarorðunum en kunnátta á markorðum sem kennd voru með beinni kennslu jókst meira en á orðum sem ekki var unnið með. Mánuði eftir að íhlutun lauk var Eva aftur prófuð úr markorðunum og hafði orðaþekkingin ekki haldist að fullu en var þó mun betri en fyrir íhlutun.



Mynd 3. Fjöldi réttra svara hjá Evu á markorðaforðaprófi fyrir íhlutun, eftir íhlutun og mánuði eftir íhlutun.

Daglegar mælingar

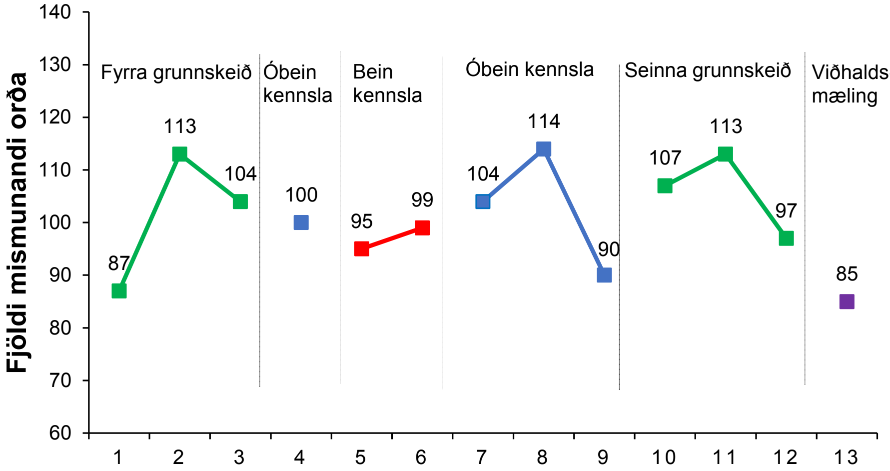
Á *Mynd 4* má sjá niðurstöður daglegu mælinganna á þekkingu á markorðum, en þau voru sex talsins. Prófað var annaðhvort úr markorðunum sem voru ýmist útskýrð sérstaklega við beina kennslu eða ekki útskýrð við óbeina kennslu. Hjá Tómasi var grunnskeið fyrstu vikuna (á 1.–3. degi) og var þá sögubók lesin án útskýringar. Hann gat einu sinni svarað rétt til um eitt orð og fékk fyrir það tvö stig. Í annarri og þriðju viku (á 4.–9. degi) hófst svo íhlutun þar sem farið var á markvissan hátt að kenna sex orð daglega með beinni kennslu. Á myndinni má sjá miklar framfarir. Í áttundu mælingu má sjá að Tómas náði að svara spurningum um öll markorðin rétt og fékk þá 12 stig. Í fjórðu til sjöttu viku (á 10.–18. degi) fór aftur fram óbein kennsla og orðin ekki útskýrð sérstaklega. Þar má sjá að færnin féll aftur niður. Hjá Evu voru grunnskeiðsmælingar teknar fyrstu fjórar vikurnar (12 daga). Þá var lesin sögubók án útskýringa, þ.e. óbein kennsla. Eins og sjá má á *Mynd 4* fékk Eva 0-4 stig fyrir rétt svör í óbeinni kennslu á grunnskeiði, þannig að hún gat mest svarað rétt til um tvö orð. Síðustu tvær vikurnar (á 13.–18. degi) hófst íhlutun þar sem Eva fékk beina kennslu á markorðunum. Þar jókst kunnátta hennar smám saman og fékk hún 8-9 stig fyrir rétt svör síðustu fjóra dagana.



Mynd 4. Daglegar mælingar á markorðaforðaprófi hjá Tómasi og hjá Evu.

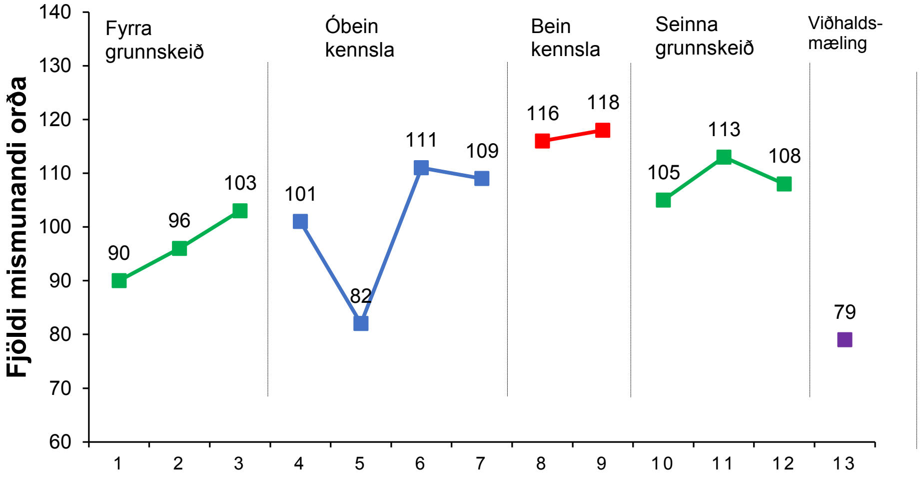
Málsýni

Samtals voru tekin 13 málsýni hjá hvoru barni. *Mynd 5* sýnir fjölda mismunandi orða (FMO) hjá Tómasi. Eins og sést voru mælingar sveiflukenndar og ekki er hægt að segja að kennslan hafi haft áhrif á hversu fjölbreytileg orð Tómas notaði í sjálfsprottnu tali. Í 1., 5., 6., 9., 12. og 13. mælingu var Tómas um einu staðalfráviki fyrir neðan meðaltal jafnaldra (Jóhanna T. Einarsdóttir og Álfhildur Þorsteinsdóttir, 2015).



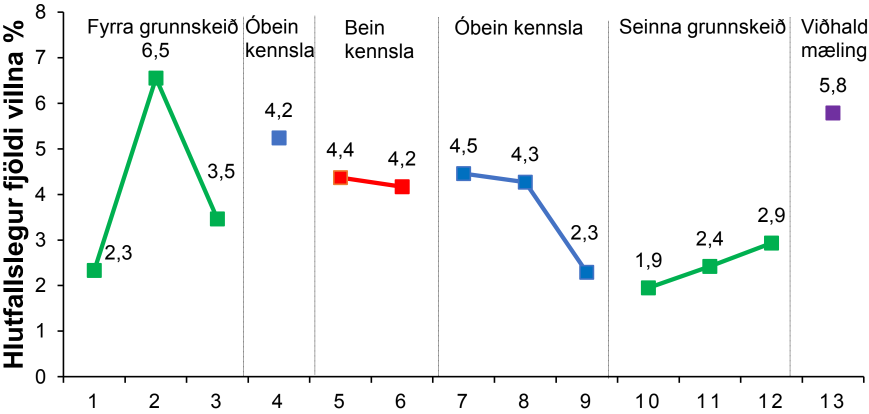
Mynd 5. Fjöldi mismunandi orða (FMO) hjá Tómasi.

Mynd 6 sýnir FMO hjá Evu. Hjá henni voru mælingar sömuleiðis nokkuð sveiflukenndar. FMO var undir meðaltali jafnaldra en þegar beina kennslan fór fram munaði einungis örfáum orðum að hún næði meðaltali barna í hennar aldurshópi (Jóhanna T. Einarsdóttir og Álfhildur Þorsteinsdóttir, 2015).



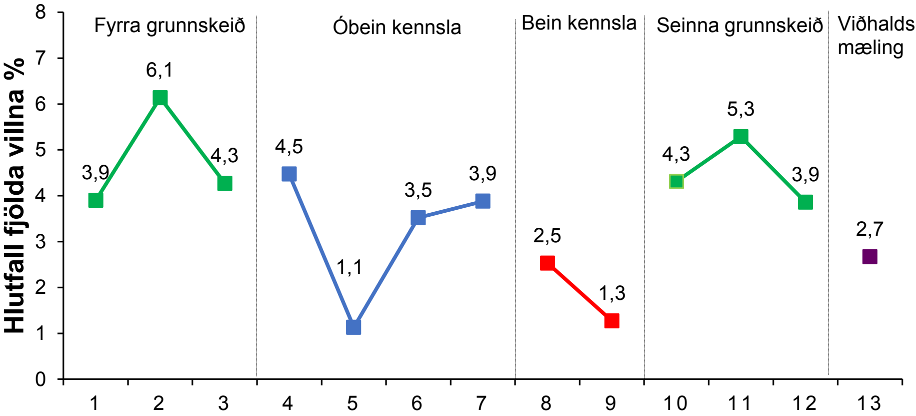
Mynd 6. Fjöldi mismunandi orða (FMO) hjá Evu.

Málfræðivillur voru mjög áberandi í tali barnanna og gerðu þau hlutfallslega mun fleiri villur en jafnaldrar að meðaltali. Dæmi um setningar sem Tómas myndaði eru: *Núna áa allir að fara að borða, þau á bara að leika sér, við á svo fullt af dýrum og þeir á að vinna í tjútjú lestinni.* Í sjálfsprottnu tali hjá jafnöldrum Tómasar voru að meðaltali 1,3% orða með villum (Jóhanna T. Einarsdóttir og Álfhildur Þorsteinsdóttir, 2015). Tómas gerði mun fleiri villur hlutfallslega í öllum mælingum. Kennslan virtist hins vegar hafa jákvæð áhrif á hlutfallslega tíðni villna á seinna grunnskeiði, en þar gerði hann færri villur í sjálfsprottnu tali. Villum fjölgaði þó aftur í viðhaldsmælingunni. Mynd 7 sýnir hlutfallslegan fjölda villna hjá Tómasi meðan á rannsókninni stóð.



Mynd 7. Hlutfallslegur fjöldi villna (%) hjá Tómasi.

Málfræðivillur voru einnig áberandi í tali Evu og sérstaklega átti hún í erfiðleikum með að beygja sagnir rétt. Þær voru oft óbeygðar og hafðar í nafnhætti eða rangt beygðar. Sem dæmi heyrðust setningar líkt og *litla barnið sofa hér, ég má ekki brotna rúmið hennar.* Einnig vantaði oft orð inn í setningar eða bygging setninganna var ófullkomin, samanber *þeir alltaf að vera í Star Wars, hann er litinn eins og þetta.* Mikil sveifla var í villufjölda á milli vikna hjá Evu. Kennslan virtist hafa jákvæð áhrif þannig að villur urðu þá færri í sjálfsprottnu tali. Hjá jafnöldrum Evu er hlutfallslegur fjöldi villna 1,6% orða (Jóhanna T. Einarsdóttir og Álfhildur Þorsteinsdóttir, 2015) og var Eva yfir því meðaltali í öllum mælingunum nema tveimur, þ.e. 5. og 9. mælingu. Mynd 8 sýnir hlutfall fjölda villna í málsýnum hjá Evu.



Mynd 8. Hlutfallslegur fjöldi villna (%) hjá Evu.

Meðferðarheldni

Þegar horft er til mælinga á magni íhlutunar var upphaflega gert ráð fyrir 24 kennslutímum (fjöldi skipta) á hvort barn eða samtals 48 tímum. Vegna veikinda hjá börnunum náðist ekki að nota alla tímana en samtals var kennslutíminn 85,4% af ráðgerðum heildarkennslutíma hjá báðum börnum. Áætluð tímalengd með beinni kennslu var 30 mínútur en 15 mínútur þegar óbein kennsla fór fram (lengd hvers skiptis). Meðaltal heildartíma með beinni kennslu var 26,8 mínútur en 8,9 mínútur með óbeinni kennslu. Heildaríhlutunartími var 25 dagar eða 6 vikur. Niðurstöður fyrir fastheldni sýndu að í 100% tilfella var hún viðunandi, það sama átti við um gæði rannsóknarinnar en þau voru í 100% tilfella viðunandi. Niðurstöður fyrir svörun sýndu að í 70,7% tilfella höfðu börnin ánægju af lestrinum en þau hlustuðu með hálfum huga í 29,3% tilfella.

Munur á mælingum rannsakanda og hins óháða rannsóknaraðila var 0,85 mínútur eða 97% samræmi í athugunum á heildartíma kennslu. Í beinu kennslunni var munur upp á 0,11 mínútur eða 99% samræmi fyrir heildartíma kennslunnar. Mælingar á fastheldni og gæðum rannsóknarinnar sýndu 100% samræmi á milli matsmanna. Mælingar á svörun þátttakenda sýndu samræmi matsmanna upp á 95,7% bæði fyrir ánægju af lestrinum og fyrir að hlusta með hálfum huga. Af þessum niðurstöðum má því álykta að rannsakandinn hafi framkvæmt rannsóknina eins og upphaflega var lagt upp með jafnvel þó heildartími kennslu væri um 85% af því sem upphaflega var gert ráð fyrir.

Umræða

Helstu niðurstöður rannsóknarinnar voru þær að með því að lesa sögubók fyrir börnin tvö og kenna markorð með beinni kennslu ásamt kennslu að sögunni lokinni náðist góður árangur og þekking barnanna á markorðunum jókst umfram þekkingu þeirra á samanburðarorðum. Lestur sögubóka án þess að útskýra markorðin sérstaklega hafði ekki sömu áhrif. Sú aukning á orðaforða sem átti sér stað í íhlutun hélst ekki að fullu mánuði eftir að henni lauk. Rannsóknin sýnir að við lestur sögubóka er mikilvægt að ræða um orð sem koma fyrir í textanum og útskýra þau. Orðaforðakennslan hafði ekki afgerandi áhrif á máltjáningu barnanna í sjálfsprottnu tali eins og hún var mæld með málsýnum en vísbendingar voru um að málfræðivillum færi fækkandi þrátt fyrir að börnin tjáðu sig með jafn mörgum mismunandi orðum.

Bein kennsla á markorðum

Fyrsta rannsóknarspurningin fjallaði um hvort Tómas og Eva myndu læra fleiri markorð en samanburðarorð. Það var metið með sérhönnuðu mælitæki sem var búið til fyrir þessa rannsókn og kannaði þekkingu barnanna á þessum orðum.

Prófun, sem var gerð fyrir og strax eftir íhlutun, staðfesti jákvæð áhrif beinnar kennslu. Tómas jók þekkingu sína á markorðunum sem voru útskýrð beint um 36,1% en þekking hans á samanburðarorðunum jókst aðeins lítillega eða um 8,3%. Eva náði einnig betur að tileinka sér markorðin sem voru kennd með beinni kennslu eða um 58,3% við íhlutun samanborið við 22,2% við óbeina kennslu. Daglegar mælingar sýndu jákvæðari áhrif beinnar kennslu á orðaforða en óbeinnar kennslu. Nokkur atriði kunna að skýra þessi áhrif beinnar kennslu. Í fyrsta lagi voru orðin endurtekin og unnið með þau sérstaklega, í öðru lagi voru þau útskýrð beint og þar með lærðu þátttakendur hvernig þeir gætu útskýrt orðin og í þriðja lagi tók beina kennslan um þrisvar sinnum lengri tíma en óbeina kennslan. Niðurstöður erlendra rannsókna eru hliðstæðar og sýna að bein kennsla hefur jákvæðari áhrif á orðanám en óbein (Beck o.fl., 2013; Biemeller og Boote, 2006; Coyne o.fl., 2007) . Í skýrslu National Institute of Child Health and Human Development (2000) var m.a. nefnt að endurtekning og að kenna orð í fjölbreyttu samhengi væri árangursrík leið til að auka orðaforða barna. Samræmist það aðferð beinnar kennslu við lestur sögubóka.

Þó að Marulis og Neuman (2010) hafi ekki getað sýnt fram á að lengd kennslutíma skipti höfuðmáli þegar verið er að auka orðaforða hjá ungum börnum er vel hugsanlegt að börn með málþroskaröskun þurfi lengri kennslutíma. Riches o.fl. (2005) komust einmitt að því að börn með sértæka málþroskaröskun þurftu að heyra nýtt orð oftar og yfir lengri tíma en samanburðarhópur sem var ekki með málþroskaröskun. Raunveruleg tímalengd íhlutunar, eins og hún var mæld í þessari rannsókn með ítarlegum mælingum á meðferðarheldni, er mjög mikilvægur þáttur og oft vanmetinn í rannsóknum (Liebermann-Betz, 2015).

Hvað varðar áhrif beinnar kennslu eru niðurstöður í takt við erlendar rannsóknir, t.d. Biemiller og Boote (2006), Coyne og félaga (2007), Marulis og Neuman (2010) og Nash og Snowling (2006) um að börn læri betur ný orð með því að fá þau útskýrð beint. Sömuleiðis hafa innlendar rannsóknir, þar sem þátttakendur voru tvítyngd börn, sýnt hliðstæðar niðurstöður (Eyrún Rakel Agnarsdóttir, 2018; Helga Hilmarsdóttir, 2016).

Rannsóknir hafa einnig sýnt að það að lesa fyrir börn hefur jákvæð áhrif á málþroska þeirra (Amalía Björnsdóttir o.fl., 2003; Hrafnhildur Ragnarsdóttir, 2015). Þó má draga þá ályktun að ekki sé sama hvernig lesið er fyrir börn, og þá sérstaklega börn með málþroskaröskun. Börn með málþroskaröskun þurfa til dæmis að heyra ný orð oftar og yfir lengri tíma en börn með dæmigerðan málþroska (Riches o.fl., 2005). Við athugun rannsakanda fyrir íhlutun, þegar fylgst var með börnunum í sögustund, kom í ljós að Tómas og Eva virtust ekki fylgjast með sem gefur til kynna að mögulega séu þau ekki að nýta sér sögulestur í hóp til að efla og auka málfærni sína, líkt og jafnaldrar með eðlilegan málþroska gera alla jafna. Í þessari rannsókn kom í ljós að þau áttu erfitt með að geta sér til um merkingu orðanna út frá samhengi textans eða að nýta sér hliðstæð orð til að læra orð sem þau kunnu ekki fyrir. Hugsanlega voru bækurnar of þungar fyrir þau eða fyrst þurfti að auka orðaforða þeirra í fyrsta lagi orðaforðans (Beck o.fl., 2013). Það er því ljóst að þau þurfa aðstoð við að læra ný orð og einnig við að skilja sögu þegar hún er lesin fyrir þau.

Viðhelst þekking á nýjum orðum?

Önnur rannsóknarspurningin sneri að því hvort Tómas og Eva næðu að viðhalda þekkingu sinni á nýlærðum orðum mánuði eftir að íhlutun lyki. Niðurstöður sýndu að þekkingin viðhélst ekki að fullu hjá börnunum. Hjá Tómasi minnkaði þekking á markorðum um 13,9% og á samanburðarorðum um 8,3% frá íhlutunarskeiði. Hjá Evu varð líka lækkun í stigafjölda á orðaforðaprófinu, og meiri á markorðum en samanburðarorðum. Lækkunin á stigafjölda markorðanna hjá Evu nam um 16,6% en um 2,8% á samanburðarorðunum mánuði eftir að íhlutun lauk. Þessar niðurstöður eru ekki í samræmi við rannsókn Coyne og félaga (2007), en þar viðhélst kunnátta barnanna á orðaforða sem kenndur var með beinni kennslu yfir átta vikna tímabil. Í meistaraprófsritgerð Helgu Hilmarsdóttur (2016), þar sem hún prófaði sams konar aðferð á tvítyngdu barni, viðhélst þekkingin á markorðunum mánuði eftir að íhlutun lauk. Hins vegar eru þessar niðurstöður í samræmi við niðurstöður rannsókna um að börnum með málþroskaröskun er hættara við að viðhalda ekki að fullu færni sem þau hafa náð með íhlutun (Riches o.fl., 2005). Tómas og Eva sýndu bæði meiri þekkingu á markorðum mánuði eftir að kennslu lauk en á prófunum fyrir íhlutun. Það má því draga þá ályktun að börn með málþroskaröskun þurfi viðhaldskennslu til að kunnátta þeirra haldist. Hugsanlega væri nóg fyrir þau að heyra bækurnar lesnar aftur til að rifja upp orðin, þó svo að orðin væru ekki útskýrð sérstaklega. Þá heyrðu þau orðin reglulega, fengju endurtekningu og heyrðu þau í samhengi við texta.

Málsýni

Þriðja rannsóknarspurningin fjallaði um hvort kennslan hefði áhrif á almenna máltjáningu í sjálfsprottnu tali eins og hún birtist í málsýnum. Þeir mælikvarðar sem notaðir voru til að meta máltjáningu barnanna voru annars vegar fjöldi mismunandi orða og hins vegar hlutfallslegur fjöldi villna.

Samkvæmt greiningu á málsýnunum virtist orðaforðaþjálfunin ekki hafa áhrif á almenna máltjáningu Tómasar og Evu í sjálfsprottnu tali. Það er að segja, fjöldi mismunandi orða jókst ekki, sem segir að orðaforðinn varð ekki fjölbreyttari þegar þau tjáðu sig í sjálfsprottnu tali. Í flestum mælingum var FMO undir meðaltali jafnaldra (Jóhanna T. Einarsdóttir og Álfhildur Þorsteinsdóttir, 2015).

Hins vegar fækkaði villum í sjálfsprottnu tali eftir því sem leið á þjálfunartímabilið þó að heildarfjöldi villna hafi í flestum mælingum verið mun meiri en hjá jafnöldrum (Jóhanna T. Einarsdóttir og Álfhildur Þorsteinsdóttir, 2015). Hlutfall fjölda villna var mest í upphafi en lækkaði eftir að kennsla hófst og virtist ekki skipta máli hvort þjálfað var með beinni eða óbeinni kennslu. Af þessu má draga þá ályktun að allt þjálfunartímabilið, þ.e. bæði þegar bein og óbein kennsla fór fram, hafi haft jákvæð áhrif þannig að villum fækkaði. Þó svo að markmiðið með þjálfuninni hafi ekki verið að fækka villum var rannsakandi sífellt að endurtaka á leiðréttandi hátt rangt tal (e. recasting). Þegar börnin sögðu setningu eða orð sem voru málfræðilega röng endurtók rannsakandi setninguna eða orðið á leiðréttandi hátt. Bæði Tómas og Eva heyrðu því stöðugt endurtekið það sem þau sögðu með réttri málfræði. Auk þess fengu þau aukna athygli og gæðastund með einum fullorðnum einstaklingi með auknum samskiptum og samtali sem kann að hafa haft jákvæð áhrif á málfærni þeirra.

Takmarkanir rannsóknarinnar

Rannsóknin var gerð að erlendri fyrirmynd en fjölmargar rannsóknir hafa sýnt fram á jákvæð áhrif beinnar kennslu með sögulestri (Beck o.fl., 2013; Biemeller og Boote, 2006; Coyne o.fl., 2007). Prófun úr markorðaforðanum þarf að taka með fyrirvara. Mælitækið var hannað af rannsakanda að erlendri fyrirmynd. Það var því hvorki um neina staðla að ræða né reglur um fyrirgjöf eða forprófun. Það var alltaf mat rannsakanda hvort þátttakendur fengu rétt eða rangt fyrir svarið. Það hefði verið ákjósanlegt að nota staðlað orðaforðapróf samhliða sérhannaða mælitækinu. Mæling með sérhönnuðu orðaforðaprófi getur gefið ýkta mynd af orðaforðaaukningu þar sem eingöngu er spurt um orðin sem verið er að vinna með. Ekki var lagt fyrir annað orðaforðapróf þar sem ekki eru til íslenskir staðlar eða viðmið fyrir þess konar próf. Hins vegar voru tekin málsýni sem gáfu góða mynd af sjálfsprottnu tali barnanna. Þar eru til íslensk viðmið sem eru afar gagnleg til að meta málþroska barna í sjálfsprottnu tali samanborið við jafnaldra (Jóhanna T. Einarsdóttir og Álfhildur Þorsteinsdóttir, 2015).

Við val á þátttakendum má horfa til þess að Tómas er alinn upp við tvö tungumál þó svo íslenska sé hans sterkara mál og það tungumál sem hann notar í samskiptum við aðra. Ákjósanlegast hefði verið að velja þátttakendur sem ólust eingöngu upp í íslensku málumhverfi, en þar sem vitað var fyrirfram að Tómas var tæpum tveimur staðalfrávikum undir meðalgetu jafnaldra á íslensku og að hann talaði litla spænsku, þótti hann henta vel í þessa rannsókn. Tómas féll því undir skilgreiningu málþroskaröskunar þar sem mikil frávik voru á báðum tungumálum í hans umhverfi.

Þegar einliðasnið er notað, snið margþætts grunnskeiðs, skal ávallt hafa í huga áhrifaþætti sem geta skekkt niðurstöður (Guðrún Árnadóttir og Þorlákur Karlsson, 2003). Í þessari rannsókn gæti það t.d. hafa verið önnur kennsla samhliða rannsókninni. Þar sem orðaforði er breyta sem erfitt er að hafa stjórn á var erfitt að útiloka ýmsa þætti sem gætu haft áhrif. Það var þó reynt eftir bestu getu, til dæmis með nákvæmum mælingum á meðferðarheldni íhlutunarinnar. Einnig með því að fjarlægja þær bækur úr leikskólanum sem notaðar voru í kennslunni. Eins tóku þátttakendur ekki þátt í skipulögðum málörvunartímum í leikskólanum og tímum í talþjálfun var hætt á meðan rannsóknin fór fram. Samt sem áður var ekki hægt að hafa stjórn á öllum öðrum áhrifaþáttum. Má þar til dæmis nefna að Tómas og Eva fóru í samverustundir á hverjum degi þar sem lesnar voru sögubækur. Ekki er hægt að útiloka að eitthvað af markorðunum 18 sem kennd voru og unnið með hafi komið fyrir í þeim bókum.

Ályktanir

Í þessari rannsókn var orðaforðakennsla könnuð með lestri sögubóka hjá tveimur börnum með málþroskaröskun. Niðurstöður benda til þess að orðaforðakennsla af þessu tagi, þar sem orð eru útskýrð um leið og þau koma fyrir í sögubók, skili betri árangri en lestur sögubóka án þess að útskýra ákveðin orð. Þessar niðurstöður eru í samræmi við erlendar rannsóknir (t.d. Biemiller og Boote, 2006 og Coyne o.fl., 2007). Hins vegar hélst árangurinn ekki að fullu mánuði eftir að íhlutun lauk. Það bendir til þess að börn með málþroskaröskun þurfi stöðuga örvun til að viðhalda orðaforða sínum. Þetta er mikilvægt að hafa í huga þegar verið er að leiðbeina foreldrum og leikskólakennurum barna sem eru með frávik í máli. Þau þurfa aðstoð við að skilja sögurnar, þau þurfa beina útskýringu á orðum og það verður að gæta þess að málfar bókanna sé við hæfi. Orðaforðakennsla af þessu tagi er mikilvæg fyrir börn með málþroskafrávik en það er ljóst að börnin þurfa meiri og markvissari örvun bæði starfsmanna leikskólans og foreldra til að málörvunin yfirfærist á sjálfsprottið tal.

Vocabulary intervention through storybook reading for children with developmental language disorder

Children diagnosed with developmental language disorders (DLD) have difficulty learning language. This affects both language comprehension and expression and occurs without any obvious explanation. Many children with DLD have coexisting conditions, such as attentional or emotional problems. Children with DLD can have different types of difficulties learning language and the severity varies. The symptoms are numerous, including problems with learning and applying the rules of grammar, sentence construction, and language use. One common symptom is limited vocabulary. Children with DLD have smaller receptive and expressive vocabularies than their peers. They learn new words at slower rate and forget newly acquired words more rapidly. They know fewer words and have weaker semantic connections within their lexical system. This deficit is visible in their expressive language as they often use simple, high frequency vocabulary. This lack of vocabulary diversity can affect their future reading comprehension and academic performance. Research have shown that DLD is common, with approximately 9% of children displaying signs of language impairment without other coexisting difficulties. Considered in the context of the population of Iceland, this means that approximately 400 children in every year-based age group could be affected by DLD.

The aim of this project was to examine the effect of an intervention which aimed to increase the vocabulary of two children diagnosed with DLD. Both attended the same preschool and were in their last year in the preschool where the training took place. The training involved reading a story book where two different methods of teaching target words were compared. The words were either (a) explained explicitly and directly when they occurred in the text, or (b) indirectly when the children were exposed to the words in the text but without explicit teaching. A multiple baseline design was employed by comparing the intervention methods between the children. The intervention took place four times a week for six weeks. Child A received indirect teaching in the first week, then two weeks of direct teaching, followed by three weeks of indirect teaching. As regards Child B, indirect teaching occurred in the first four weeks, followed by direct teaching during the last two weeks. The effect of the intervention was measured by testing how well the children could define the target words and by language samples of spontaneous speech. Measurements were made before, during, and immediately following training, as well as a month after training ended.

Expanding vocabulary by direct teaching was found to be a more powerful method than indirect teaching. Vocabulary training using direct teaching had a positive effect with regard to the words taught. Measurements showed that when direct teaching was conducted the children in this study knew the words and were able to explain

them, whereas their vocabulary improved only slightly when indirect teaching methods were applied. Measurements taken a month after intervention showed that the children did not fully retain the newly gained vocabulary. It is clear, however, that to maintain the newly gained vocabulary repeated reading, including discussion of target words, is necessary for children with DLD. Reading storybooks with or without direct teaching of target words did not affect the children's spontaneous vocabulary use as measured by language samples; that is, they did not begin to use more complex and different words when speaking spontaneously. However, on average the children produced more grammatically correct sentences and made fewer morphemic errors.

The findings of this study show the importance of reading story books to children during their preschool years. The findings demonstrate that for learning new words children with poor language skills need direct teaching and repeated intensive instruction. It is not enough merely to read the stories without explaining unknown new words. Furthermore, repeated measurements showed that during and after the intervention the children spoke with fewer grammatical errors. This is probably because the researcher made intensive use of the method of recasting during the intervention. Recasting has been shown to be an effective way of correcting syntactic and grammatical errors.

The study also demonstrated the importance of supporting children with poor language skills and facilitating their language acquisition. These results can be used for the benefit of parents, teachers and speech pathologists alike.

Key words: Language development, developmental language disorder, story book reading, explicit and implicit vocabulary intervention

Um höfundana

Sigrún Alda Sigfúsdóttir (sigrunalda88@gmail.com) er talmeinafræðingur og starfar á Talsetrinu og hjá Reykjavíkurborg. Sigrún Alda lauk MS-prófi í talmeinafræði frá Háskóla Íslands árið 2018 og BA-prófi í félagsráðgjöf frá Háskóla Íslands árið 2012.

Jóhanna T. Einarsdóttir (jeinars@hi.is) er talmeinafræðingur og dósent við Háskóla Íslands á Mennta- og Heilbrigðisvísindasviði. Hún lauk doktorsprófi frá Læknadeild Háskóla Íslands árið 2009. Rannsóknir Jóhönnu Thelmu hafa beinst að máltöku barna á leikskólaaldri, bæði mælingum á málþroska og íhlutun. Hún hefur einnig rannsakað stam, greiningu þess og meðferð hjá öllum aldurshópum.

Þorlákur Karlsson (thorlakur@ru.is) er sálfræðingur og dósent í sálfræði við Háskólann í Reykjavík. Hann er með doktorspróf í sálfræði frá West Virginia University (1990) með atferlisgreiningu sem aðalgrein en tölfræði og aðferðafræði sem aukagrein. Rannsóknir hans eru helst á sviði aðferðafræði rannsókna og launamunar kynja.

Íris Ösp Bergþórsdóttir (iob@hi.is) stundar doktorsnám við Heilbrigðisvísindasvið Háskóla Íslands, rannsókn hennar snýr að meðferðarheldni í foreldramiðaðri stamrannsókn á börnum. Hún lauk MS-prófi í Viðskiptafræði frá Háskóla Íslands árið 2006 og BS-prófi í sálfræði frá Háskóla Íslands árið 2001. Rannsóknir hennar eru helst á sviði meðferðarheldnis og frammistöðumatsmælinga.

About the authors

Sigrún Alda Sigfúsdóttir (sigrunalda88@gmail.com) is a speech-language pathologist working for Talsetrið and The City of Reykjavík. Sigrún Alda completed an M.S. degree in speech-language pathology from the medical department at the University of Iceland in 2018 and a bachelor's degree in Social studies from the University of Iceland in 2012.

Jóhanna T. Einarsdóttir (jeinars@hi.is) is an associate professor at the University of Iceland School of Education and School of Health Sciences. She received her PhD in 2009 from the medical department at the University of Iceland. Jóhanna has been investigating children's language acquisition in early childhood, measurements as well as intervention. Her special interests are fluency disorders, stuttering and stuttering treatment.

Thorlakur Karlsson (thorlakur@ru.is) is a psychologist and an associate professor at Reykjavík University. He has a doctoral degree in psychology from West Virginia University (1990), with a major in behavior analysis and a minor in statistics and methodology. His research is mainly in the area of methodology and gender wage difference.

Íris Ösp Bergþórsdóttir (iob@hi.is) is a PhD student within the School of Health Science at the University of Iceland. Her doctoral research investigates implementation fidelity in parent-managed treatments for children. Íris completed her M.S. degree in business from the University of Iceland in 2006 and a bachelor's degree in psychology from the University of Iceland in 2001. Her research is mainly in the area of fidelity and performance research.

Heimildir

Amalía Björnsdóttir, Ingibjörg Símonardóttir og Jóhanna Einarsdóttir. (2003). Þróun HLJÓM-2 og tengsl þess við lestrarfærni og ýmsa félagslega þætti. *Uppeldi og menntun, 12,* 9–30.

Árdís Hrönn Jónsdóttir. (2013). *Orðaspjall: Að efla orðaforða og hlustunarskilning barna með bókalestri*. Reykjanesbær: Leikskólinn Tjarnarsel.

Beach, K. D., Sanchez, V., Flynn, L. J. og O'Connor, R. (2015). Teaching academic vocabulary to adolescents with learning disabilities. *Teaching Exceptional Children, 48*(1), 36–44. doi:10.1177/0040059915594783

Beck, I. L. og McKeown, M. G. (2007). Increasing young low-income children's oral vocabulary repertoires through rich and focused instruction. *Elementary School Journal, 107*(3), 251–271. doi:10.1086/511706

Beck, I. L., McKeown, M. G. og Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction*. New York, NY: Guilford Press.

Beck, I. L., Perfetti, C. A. og McKeown, M. G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology, 74*(4), 506–521. doi:10.1037/0022-0663.74.4.506

Biemiller, A. og Boote, C. (2006). An effective method for building meaning vocabulary in primary grades. *Journal of Educational Psychology, 98*(1), 44–62. doi:10.1037/0022-0663.98.1.44

Bishop, D. V. M., Snowling, M. J., Thompson, P. A. og Greenhalgh, T. (2016). CATALISE: A multinational and multidisciplinary Delphi consensus study. Identifying language impairments in children. *PLoS ONE 11*(7), 1–26. doi:10.1371/journal.pone.0158753

Bishop, D. V. M., Snowling, M. J., Thompson, P. A. og Greenhalgh, T. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry, 58*(10), 1068–1080. doi:10.1111/jcpp.12721

Catts, H.W., Fey, M. E., Tomblin, J. B. og Zhang, X. (2002). A longitudinal investigation of reading outcomes in children with language impairments. *Journal of Speech, Language, and Hearing Research, 45*(6), 1142–1157. doi:10.1044/1092-4388(2002/093)

Coyne, M. D., McCoach, D. B. og Kapp, S. (2007). Vocabulary intervention for kindergarten students: Comparing extended instruction to embedded instruction and incidental exposure. *Learning Disability Quarterly, 30*(2), 74–88. doi:10.2307/30035543

Dusenbury, L., Brannigan, R., Falco, M. og Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research, 18*(2), 237–256. doi:10.1093/her/18.2.237

Eyrún Rakel Agnarsdóttir. (2018). *Þjálfun orðaforða tvítyngdra barna með lestri sögubóka. Bein og óbein þjálfun* (óútgefin meistararitgerð). Heilbrigðisvísindasvið, Háskóli Íslands, Reykjavík.

Guðrún Árnadóttir og Þorlákur Karlsson. (2003). Einliðasnið: Öflug leið til samhæfingar klínískrar vinnu og rannsókna. Í Sigríður Halldórsdóttir og Kristján Kristjánsson (ritstjórar), *Handbók í aðferðafræði og rannsóknum í heilbrigðisvísindum* (bls. 295–330). Akureyri: Háskólinn á Akureyri.

Helga Hilmarsdóttir. (2016). *Orðaforði tvítyngdra barna – Orðaforðaþjálfun með lestri sögubóka* (óútgefin meistararitgerð). Heilbrigðisvísindasvið, Háskóli Íslands, Reykjavík.

Hiebert, E. H. og Kamil, M. L. (2005). *Teaching and learning vocabulary: Bringing research to practice.* London, UK: Routledge. doi:10.4324/978141061

Hrafnhildur Ragnarsdóttir. (2015). Málþroski leikskólabarna. Þróun orðaforða, málfræði og hlustunarskilnings milli fjögra og fimm ára aldurs. *Netla – Veftímarit um uppeldi og menntun.* Sótt af http://netla.hi.is/greinar/2015/ryn/007.pdf

Jóhanna T. Einarsdóttir, Amalía Björnsdóttir og Ingibjörg Símonardóttir. (2016). The predictive value of preschool language assessments on academic achievement: A 10-year longitudinal study of Icelandic children. *American Journal of Speech-Language Pathology, 25*(1), 67–79. doi:10.1044/2015_AJSLP-14-0184

Jóhanna T. Einarsdóttir og Álfhildur Þorsteinsdóttir. (2015). Málsýni leikskólabarna: Aldursbundin viðmið. *Netla – Veftímarit um uppeldi og menntun.* Sótt af http://netla.hi.is/greinar/2015/ryn/010.pdf

Jóhanna Thelma Einarsdóttir og Stefán Carl Peiser. (2016). *Málgreinir.* Sótt af http://malsyni.hi.is/?page_id=184

Jóhanna Thelma Einarsdóttir og Þóra Sæunn Úlfsdóttir. (2018). *Málsýnataka.* Sótt af http://malsyni.hi.is/wp-content/uploads/2018/02/handb%C3%B3k-6.2.2018.pdf

Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings* (2. útgáfa). Oxford, UK: Oxford University Press.

Laufer, B. og Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language, 22*(1), 15–30.

Law, J. (2019). Introduction. Í J. Law, C. McKean, C. A. Murphy og E. Thordardottir (ritstjórar), *Managing children with developmental language disorders. Theory and practice across Europe and beyond* (bls. 3–5). London, UK: Routledge.

Liebermann-Betz, R. G. (2015). A systematic review of fidelity of implementation in parent-mediated early communication intervention. *Topics in Early Childhood Special Education, 35*(1), 15–27. doi:10.1177/0271121414557282

Marulis, L. M. og Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research, 80*(3), 300–334. doi:10.3102/0034654310377087

McGregor, K. K., Oleson, J., Bahnsen, A. og Duff, D. (2013). Children with developmental language impairment have vocabulary deficits characterized by limited breadth and depth. *International Journal of Language and Communication Disorders, 43*(3), 307–319. doi:10.1111/1460-6984.12008

McKeown, M. G. og Beck, L. (2014). Effects of vocabulary instruction on measures of language processing: Comparing two approaches. *Early Childhood Research Quarterly, 29*(4), 520–530. doi:10.1016/j.ecresq.2014.06.002

McKeown, M. G., Beck, I. L., Omanson, R. C. og Perfetti, C. A. (1983). *The effects of long-term vocabulary instruction on reading comprehension: A replication. Journal of Reading Behavior, 15*(1), 3–18. doi:10.1080/10862968309547474

McKeown, M. G., Beck, I. L., Omanson, R. C. og Pople, M. T. (1985). Some effects of the nature and frequency of vocabulary instruction on the knowledge and use of words. *Reading Research Quarterly, 20*(5), 522–535. doi:10.2307/747940

Nash, H. og Snowling, M. (2006). Teaching new words to children with poor existing vocabulary knowledge: A controlled evaluation of the definition and context methods. *International Journal of Language & Communication Disorders, 41*(3), 335–354. doi:10.1080/13682820600602295

Nation, K. og Snowling, M. J. (2004). Beyond phonological skills: Broader language skills contribute to the development of reading. *Journal of Research in Reading, 27*(4), 342–356. doi:10.1111/j.1467-9817.2004.00238.x

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence based assessment of the scientific research literature on reading and its implications for reading instruction.* Washington, DC: Government Printing Office.

Paul, R., Norbury, C. og Gosse, C. (2018). *Language disorders from infancy through adolescence: Assessment and intervention* (5. útgáfa). St. Louis, MA: Mosby Elsevier.

RADLD. (e.d.). *Raising awareness of developmental language disorder.* Sótt af https://radld.org/

Rice, M. L., Buhr, J. C. og Nemeth, M. (1990). Fast mapping word-learning abilities of language-delayed preschoolers. *Journal of Speech and Hearing Disorders, 55*(1), 33–42. doi:10.1044/jshd.5501.33

Riches, N. G., Tomasello, M. og Conti-Ramsden, G. (2005). Verb learning in children with SLI: Frequency and spacing effects. *Journal of Speech, Language and Hearing Research, 48*(6), 1397–1411. doi:10.1044/1092-4388(2005/097)

Walsh, B. A. og Blewitt, P. (2006). The effect of questioning style during storybook reading on novel vocabulary acquisition of preschoolers. *Early Childhood Education Journal, 33*(4), 273–278. doi:10.1007/s10643-005-0052-0

Wasik, B. A. og Bond, M. A. (2001). Beyond the pages of a book: Interactive book reading and language development in preschool classrooms. *Journal of Educational Psychology, 93*(2), 243–250. doi:10.1037/0022-0663.93.2.243

Wilkinson, K. S. og Houston-Price, C. (2013). Once upon a time, there was a pulchritudinous princess... : The role of word definitions and multiple story contexts in children's learning of difficult vocabulary. *Applied Psycholinguistics, 34*(3), 591–613. doi:10.1017/S0142716411000889

Sigrún Alda Sigfúsdóttir, Jóhanna T. Einarsdóttir, Þorlákur Karlsson og Íris Ösp Bergþórsdóttir. (2020).
Orðaforðakennsla með sögulestri fyrir börn með málþroskaröskun
Netla – Veftímarit um uppeldi og menntun. Menntavísindasvið Háskóla Íslands.
Sótt af http://netla.hi.is/greinar/2020/ryn/04
DOI: https://doi.org/10.24270/netla.2020.4

Appendix B

Tvítyngd börn, orðaforði og sögulestur

Jóhanna Thelma Einarsdóttir, Eyrún
Rakel Agnarsdóttir, Íris Ösp Bergþórsdóttir

Inngangur

Fjöldi leikskólabarna á Íslandi sem tala annað tungumál en íslensku heima fyrir hefur margfaldast á síðustu tveimur áratugum. Árið 2020 voru þau um 15,6% en árið 1998 um 3,8% (Hagstofa Íslands, 2022a, 2022b). Á Íslandi sækja um 98% barna leikskóla frá unga aldri og langflest (96%) dvelja þar sjö tíma á dag eða lengur (Hagstofa Íslands, 2022a). Málörvun barna er snar þáttur í starfi leikskólans og hún er sérstaklega dýrmæt fyrir börn sem hafa annað tungumál en íslensku að heimamáli.

Rannsóknir hafa sýnt að bæði íslenskur orðaforði (Aneta Figlarska o.fl., 2017) og hljóðkerfisvitund (Kriselle Lou Susan Jónsdóttir o.fl., 2018) leikskólabarna með önnur móðurmál en íslensku er marktækt slakari en hjá dæmigerðum eintyngdum börnum. Enn fremur sýndu nýlegar niðurstöður Hjördísar Hafsteinsdóttur o.fl. (2022) að heildarmálþroski, bæði málskilningur og máltjáning, 25 tvítyngdra barna var um þremur staðalfrávikum frá meðalgetu eintyngdra jafnaldra á staðlaða málþroskaprófinu MELB (Þóra Másdóttir o.fl., 2021). Jafnframt sýndi athugun á sjálfsprottnu tali barnanna að þau tjáðu sig í marktækt styttri setningum, notuðu ekki eins fjölbreytt orð og gerðu hlutfallslega fleiri villur samanborið við eintyngda jafnaldra. Svipaða sögu er að segja um grunnskólanemendur með annað móðurmál en íslensku. Íslenskukunnátta þeirra er mun slakari í samanburði við eintyngda jafnaldra og framfarir í íslensku hægari (Elín Þöll Þórðardóttir, 2021; Elín Þöll Þórðardóttir og Anna Guðrún Júlíusdóttir, 2013; Sigríður Ólafsdóttir, 2015; Sigríður Ólafsdóttir o.fl., 2016). Rannsóknir hafa sýnt fram á mikilvægi þess að hafa náð góðum tökum á tungumálinu við lok leikskóla þar sem sterk fylgni er á milli færni í íslensku á leikskólaaldri, síðara námsgengi (Jóhanna T. Einarsdóttir o.fl., 2016) og almennri líðan í grunnskólanum (Amalía Björnsdóttir o.fl., 2013). Orðaforði er sá þáttur málþroskans sem tengist hvað sterkast málumhverfi barna (Hart og Risley, 1995; Hrafnhildur Ragnarsdóttir, 2015). Að efla íslenskan orðaforða er því áhrifamikill liður í því að auka færni tvítyngdra barna í íslensku og getur skipt sköpum fyrir velgengni þeirra í íslensku samfélagi (Sigríður Ólafsdóttir o.fl., 2016).

Málþroski, orðaforði og tvítyngi

Margir þættir hafa áhrif á hvernig börn læra tungumál. Magn og gæði samskipta við barnið skipta þar sköpum, bæði í tileinkun móðurmálsins (Pearson o.fl., 2007) og í máltöku annars máls (Elín Þöll Þórðardóttir, 2011; Paradis og Jia, 2016). Með magni er átt við hversu ríkuleg málörvun er í umhverfinu og hve miklum hluta af vökutíma börnin dvelja í hvoru tungumáli fyrir sig. Sýnt hefur verið fram á að tvítyngd börn læra frekar orð og málfræði þess tungumáls sem þau dvelja lengur í (Elín Þöll Þórðardóttir, 2011; Hoff o.fl., 2012). Gæðin skipta líka máli og að um sé að ræða jákvæð tengsl milli viðmælanda og barnanna (Brooks og Kempe, 2012; Elín Þöll Þórðardóttir, 2011). Almennt er miðað við að börn með ensku sem annað mál séu fjögur til sex ár að ná samsvarandi viðmiðum í ensku máli og eintyngdir jafnaldrar (Elín Þöll Þórðardóttir, 2021). Ýmislegt bendir til þess að það taki börn sem hafa önnur heimamál en íslensku lengur en sex ár að ná viðmiðum eintyngdra jafnaldra ef þau ná þeim yfirhöfuð (Elín Þöll Þórðardóttir, 2021; Elín Þ. Þórðardóttir og Anna Guðrún Júlíusdóttir, 2013; Hjördís Hafsteinsdóttir o.fl., 2022; Sigríður Ólafsdóttir og Hrafnhildur Ragnarsdóttir, 2010). Sérstaklega þarf að huga að orðaforða tvítyngdra barna þar sem leikskólaumhverfið er yfirleitt eini möguleiki þeirra til að læra ný íslensk orð. Algeng orð lærast í samskiptum við umhverfið en fátíðari orð þarf að kenna sérstaklega (Beck o.fl., 2013; Brooks og Kempe, 2012). Hjá tvítyngdum börnum dreifist orðaforðinn á fleiri en eitt tungumál, það er að segja bæði á heimamálið og skólamálið. Sýnt hefur verið fram á að orðaforði tvítyngdra barna á skólamálinu við níu ára aldur spáir fyrir um framfarir þeirra í lesskilningi á unglingsárum (Sigríður Ólafsdóttir o.fl., 2016). Mikilvægt er því að hafa í huga að tvítyngd börn hafi bæði djúpa þekkingu á íslenskum orðum og þekki nægilega mörg orð til að spjara sig í íslensku samfélagi.

Orðaforði og þróun hans

Orðaforði er sá þáttur málþroskans sem er hvað háðastur umhverfi barnanna og skiptir þar sköpum að alast upp í málhvetjandi umhverfi (Hart og Risley, 1995; Hoff o.fl., 2012; Noble o.fl., 2005). Talið er að börn læri u.þ.b. sjö til níu orð á dag milli 18 og 24 mánaða aldurs (Brooks og Kempe, 2012). Að meðaltali tvöfaldast orðaforðinn milli tveggja og þriggja ára aldurs (Elín Þöll Þórðardóttir, 1998) og við upphaf grunnskólagöngu er orðaforði enskumælandi barna talinn vera um 4000–6000 orð (Brooks og Kempe, 2012).

Rannsóknir á orðaforða eintyngdra íslenskra leikskólabarna hafa sýnt mikinn einstaklingsmun innan barnahópsins (Hrafnhildur Ragnarsdóttir, 2015; Jóhanna T. Einarsdóttir og Álfhildur Þorsteinsdóttir, 2015). Jóhanna og Álfhildur (2015) skoðuðu máltjáningu 221 barns á aldrinum tveggja og hálfs árs til sex og hálfs árs með því að taka málsýni af sjálfsprottnu tali. Athugað var hversu langar segðir (setningar) börnin mynduðu að meðaltali, hver væri heildarfjöldi orða, fjöldi mismunandi orða og hlutfall villna. Niðurstöður leiddu í ljós að mikil dreifing var innan hvers aldurshóps. Færni sumra barna við fjögurra ára aldur var svipaður meðalgetu sex ára barna meðan önnur fjögurra ára börn tjáðu sig með svipuðum hætti og tveggja ára börn. Niðurstöður úr rannsókn Hrafnhildar Ragnarsdóttur (2015) sýndu einnig gríðarlegan einstaklingsmun hjá 111 fjögurra ára börnum. Tengdist orðaforði barnanna menntun móður, fjölskyldutekjum, fjölda barnabóka á heimilinu, lestri fyrir barnið heima og því hvort barnið bjó á einu heimili eða tveimur. Þessar niðurstöður eru hliðstæðar þeim sem komið hafa fram í erlendum rannsóknum um tengsl félagslegra þátta við orðaforða barna (Hart og Risley, 1995; Noble o.fl., 2005). Sömu félagslegu breytur virðast einnig hafa áhrif á þróun orðaforða fjöltyngdra barna á Íslandi (Sigríður Ólafsdóttir o.fl., 2016)

Orðanám og kennsla á nýjum orðum

Börn tileinka sér orð með því að nota þau orð sem þau heyra í samskiptum, þegar talað er beint til þeirra og einnig þau sem þau heyra í samræðum hjá öðrum og þegar lesnar eru fyrir þau sögubækur (Brooks og Kempe, 2012;

Grøver o.fl., 2020; Rice o.fl., 1990). Til þess að efla orðaforða barna er áhrifaríkt að kenna þeim ný orð með mismunandi hætti og á það bæði við um eintyngd og tvítyngd börn. Rannsóknir benda sterklega til þess að hægt sé að auka orðaforða barna með því að lesa fyrir þau sögubækur. Í barnabókum eru oft orð sem eru ekki algeng í daglegu tali og þarf að kenna sérstaklega. Þau koma endurtekið fyrir og í mismunandi samhengi sem getur auðveldað börnum að skilja merkingu þeirra. Myndskreyttar bækur geta auðveldað börnum að muna orðin þar sem þau geta tengt orðið við myndina (Beck o.fl., 2013). Margar erlendar rannsóknir hafa sýnt fram á árangur þess að efla orðaforða með markvissum aðferðum. Í safngrein Marulis og Neumans (2010) voru teknar saman 67 rannsóknir þar sem könnuð var orðaforðakennsla barna á leikskólaaldri. Skoðað var hvaða kennsluaðferðir voru áhrifamestar til að auka orðaforða barna. Niðurstöður sýndu að árangursríkast var að lesa með börnunum og fyrir þau, kenna þeim ákveðin markorð sem komu fyrir í bókinni, ræða orðin og útskýra sérstaklega með beinni kennslu.

Aðferðir við kennslu á nýjum orðum með lestri sögubóka

Lestur bóka gefur tækifæri á að efla málþroska barna og þá allra helst til þess að auka orðaforða þeirra. Zevenbergen og Whitehurst (2003) þróuðu aðferðina samræðulestur (e. *dialogic reading*) sem felur í sér að lesandinn (foreldri, kennari eða aðrir) les fyrir börn og hvetur þau til að taka virkan þátt með það að markmiði að auka máltjáningu og orðaforða þeirra. Þeir hafa gert grein fyrir aðferðinni og hvernig hún er byggð upp eftir aldri og þörfum þeirra barna sem lesið er fyrir hverju sinni. Lesandinn spyr börnin opinna spurninga líkt og „hvernig, hvenær og af hverju?“. Hann endurtekur svör barnanna með jákvæðu viðmóti og leiðréttir eða bætir við og fylgir þeim eftir með annarri spurningu. Með því að hrósa börnunum og hvetja þau leitast hann við að auka áhuga þeirra. Til þess að fanga athygli þeirra og auka úthald á meðan á lestrinum stendur er talið mikilvægt að lesandinn lifi sig inn í lesturinn með því t.d. að breyta röddinni og nota látbragð. Áhugi barnanna skiptir máli og því er ekki aðalatriðið að lesa hvert einasta orð og skoða allar myndirnar heldur fylgja áhuga og úthaldi barnanna. Aðferðin Orðaspjall (e. *Text Talk*) (Beck o.fl., 2013) svipar til samræðulesturs. Orðaspjall felur í sér að kennari les upphátt sögubók fyrir börn og vinnur markvisst að því að auka orðaforða þeirra. Hann velur ákveðin orð úr bókinni til að kenna og staldrar við það þegar það kemur fyrir í textanum og útskýrir merkingu þess. Börnin ræða um orðin og leika með þau að loknum lestri. Kennarinn rekur söguþráðinn með börnunum og hvetur þau til að velta honum fyrir sér til að skapa samræður. Markmiðið er að auka orðaforða með markvissum hætti og um leið að efla skilning barna á samfelldum texta. Árdís Hrönn Jónsdóttir (2013) lagaði aðferðina Orðaspjall að íslensku í leikskólanum Tjarnarseli í Reykjanesbæ og gefin hefur verið út samnefnd bók um hana.

Bein kennsla orða með sögulestri

Bein kennsla markorða felst í því að útskýra ákveðin sérvalin orð um leið og þau koma fyrir þegar lesið er fyrir börn. Er hún oft borin saman við óbeina kennslu þar sem barnið getur sér til um hvað orðið þýðir út frá samhengi í texta og án þess að það sé útskýrt sérstaklega. Við beina kennslu heyra börnin markorðin oft, sem getur hjálpað þeim að tengja merkingu við orðin og endurtekningin eykur líkur á að orðið festist frekar í minni. Ungir hlustendur spyrja oft ekki hvað ákveðið orð þýðir á meðan á sögulestri stendur, sérstaklega þegar lesið er fyrir börn í hóp. Án aðstoðar skilja þau ef til vill ekki öll orðin og geta ekki ráðið í merkingu þeirra á meðan þau hlusta á sögu (Biemiller og Boote, 2006).

Coyne o.fl. (2004) og Beck og Mckeown (2007) voru meðal fyrstu fræðimanna sem rannsökuðu orðanám leikskólabarna með því að útskýra fyrirfram ákveðin markorð. Áður höfðu m.a. Beck o.fl. (1982) og McKeown o.fl. (1985) skoðað slíka aðferð meðal grunnskólabarna og hafði hún bæði aukið orðaforða og málskilning þeirra barna sem tóku þátt. Aðferðin fólst í því að börnin fengu beina útskýringu á markorðum við lestur sögubóka ásamt því að fá að tjá sig um markorðin og reynslu sína í tengslum við þau. Rannsóknir þeirra leiddu í ljós að börn á síðasta ári í leikskóla lærðu frekar sjaldgæf flókin orð ef farin var sú leið að útskýra orðin beint. Hliðstæðar niðurstöður komu fram hjá Justice o.fl. (2005) sem sýndu fram á að börn með slakan orðaforða juku við orðaforða sinn þegar þau fengu útskýringu á nýju orði þegar þau heyrðu það í fyrsta skipti. Sömu niðurstöðu fengu Sigrún Alda Sigfúsdóttir o.fl. (2020) með athugun á orðanámi barna með málþroskaröskun (DLD) á síðasta ári í leikskóla. Fleiri fræðimenn, eins og Biemiller og Boote (2006), sem báru saman beina kennslu og óbeina með lestri sögubóka, sýndu fram á að beina kennslan gaf betri raun við að auka orðaforða barnanna. Nýleg rannsókn frá Noregi sýndi jafnframt að samræðulestur með markvissri orðaforðakennslu var árangursrík aðferð til að auka orðaforða fjöltyngdra leikskólabarna (Grøver o.fl., 2020).

Hvaða orð á að kenna?

Vandasamt getur verið að velja orð sem á að kenna til að auka orðaforða yngri barna. Hver bók getur innihaldið mörg orð sem börn þekkja ekki og því þarf að einblína á fá markorð í einu. Fyrir yngstu börnin eru tvö til þrjú orð nóg þegar tekið er tillit til athygli og minnis barnanna. Auk þess er hætta á að þau missi söguþráðinn ef lesandinn gerir oft hlé á lestrinum til að ræða og útskýra orð. Gott er að velja orð sem nýtast börnunum á einhvern hátt vel, orð sem tengjast starfi leikskólans eða lykilorð sem nauðsynlegt er fyrir barnið að kunna til að skilja efni sögunnar sem lesin er. Einnig getur verið gott að skoða hvort barnið fái tækifæri til að nota orðið við aðrar aðstæður og hægt sé að útskýra merkingu orðsins á þann hátt að hún skili sér til barnsins (Spencer o.fl., 2012). Enn fremur má virkja börnin sjálf við að velja orðin. Beck o.fl. (1987) voru meðal þeirra fyrstu sem vöktu athygli á því hvaða orð ætti að velja fyrir orðaforðakennslu barna. Þær komu fram með svokallað orðaforðalíkan (e. *word tiers*) árið 1987 en líkanið varð enn meira áberandi í bókum þeirra síðar (sjá Beck o.fl., 2008). Í líkaninu voru útskýrð mismunandi hlutverk og gagnsemi orða í tungumálinu og orðunum skipt í þrjú lög. Í fyrsta laginu (e. *tier one*) er grunnorðaforði, sem eru þau orð sem mest eru notuð í töluðu máli, t.d. *hús*, *hundur* og *hlaupa*. Í öðru lagi (e. *tier two*) eru orð sem spanna breiðara svið og eru sjaldgæfari en í fyrsta lagi. Þessi orð koma oftar fyrir í lesmáli en töluðu máli og eru jafnvel samheiti algengra orða úr fyrsta lagi, t.d. *kröftugur*, *vernda* og *hrasa*. Í þriðja lagi (e. *tier three)* eru svo sjaldgæf og sérhæfð orð eins og *málstol.* Sum orð í þriðja lagi eru mjög sjaldgæf þannig að jafnvel fullorðið fólk þekkir þau ekki. Beck o.fl. (2008) telja að í flestum tilfellum væri æskilegast að kenna orð í lagi tvö þar sem þau nýtast börnunum best í daglegu lífi. En fyrir börn með mjög takmarkaðan orðaforða getur verið nauðsynlegt að kenna orð innan fyrsta lags. Fleiri aðferðir hafa komið fram við val á orðum til þess að nota í orðaforðakennslu. Tíðni orða getur skipt máli eða hversu oft þau koma fyrir í rituðu máli. Stahl og Nagy (2006) skiptu orðum í tvo flokka, annars vegar hátíðniorð (e. *high frequency words*) og hins vegar mjög gagnleg almenn orð (e. *high utility general vocabulary*). Þeir álitu að gott væri að velja bæði gagnleg almenn orð og hátíðniorð til að nota í orðaforðakennslu, kenna orðin og þýðingu þeirra. Enn aðrir hafa beitt mjög svipuðum aðferðum og þeirri sem Beck o.fl. (1987) komu fram með á sínum tíma. Í þeirri rannsókn sem hér er kynnt var stuðst við orðaforðalíkan þeirra Becks o.fl. (2013) og orðin voru valin úr öðru lagi orðalíkansins. Mikil þörf er á íhlutunarrannsóknum meðal tvítyngdra barna á Íslandi þar sem rannsóknir hafa sýnt að þau standa höllum fæti þar í samanburði við eintyngda jafnaldra (Elín Þöll Þórðardóttir, 2021; Hjördís Hafsteinsdóttir o.fl., 2022; Sigríður Ólafsdóttir o.fl., 2016). Finna þarf gagnreyndar leiðir til að efla málfærni þeirra og er þessi rannsókn viðleitni til að bæta úr því. Sjónum var beint að því að auka orðaforða tvítyngdra barna með því að lesa fyrir þau sögu og kenna markvisst valin orð og efla þannig færni þeirra í íslensku.

Markmið

Meginmarkmið rannsóknarinnar var að bera saman beina og óbeina kennslu hjá tvítyngdum íslenskum börnum. Rannsóknarspurningar sem leita á svara við eru eftirfarandi:

1) Skilar bein kennsla á markorðum með sögulestri betri árangri við að auka orðaforða hjá tvítyngdum börnum á leikskólaaldri en óbein kennsla þar sem markorðin eru ekki útskýrð sérstaklega?
2) Hefur bein kennsla á markorðum með lestri sögubóka áhrif á almenna máltjáningu eins og hún er metin með málsýnum?

Aðferð

Þátttakendur

Þátttakendur voru tvö börn og fór rannsóknin fram í leikskóla þeirra. Börnin voru bæði valin eftir hentugleika en haft var í huga að þau hefðu dvalið í nokkur ár í leikskóla og haft tækifæri til að læra íslensku og að þau væru ekki með DLD eða málþroskaröskun. Því var miðað við að þau uppfylltu eftirfarandi skilyrði: 1) vera tvítyngd, 2) á seinasta ári í leikskóla og 3) mælast með dæmigerðan málþroska á íslensku. Þau eru kölluð Sigríður og Viktor í rannsókninni en það eru ekki rétt nöfn þeirra. Foreldrar beggja barnanna voru pólskir og höfðu búið á Íslandi í um áratug. Móðir Sigríðar talaði ágæta íslensku en faðir hennar og foreldrar Viktors töluðu nánast enga íslensku. Foreldrar beggja barna töldu að færni þeirra á pólsku væri svipuð og hjá eintyngdum pólskum jafnöldrum þeirra. Þátttakendur fæddust á Íslandi og höfðu verið í íslensku málumhverfi frá unga aldri, fyrst hjá dagmóður og síðan í leikskólum. Þau gengu í sama leikskóla en voru ekki saman á deild. Í leikskólanum töluðu þau stöku sinnum pólsku saman en annars töluðu þau íslensku. Þátttakendur voru bæði á sínu sjötta ári þegar rannsóknin hófst og var Viktor 5 ára og 10 mánaða en Sigríður 5 ára og 7 mánaða. Málþroskaprófið TOLD-2P var lagt fyrir börnin (Ingibjörg Símonardóttir o.fl., 1995) fyrir rannsókn og sýndu niðurstöður að þau bjuggu yfir meðalfærni í íslensku miðað við eintyngda íslenska jafnaldra. Foreldrar beggja barnanna fengu ítarlegar upplýsingar um rannsóknina og skrifuðu undir samþykkisyfirlýsingu um þátttöku þeirra áður en rannsókn hófst.

Mælingar

Mælingar voru gerðar fyrir íhlutun, meðan á henni stóð og strax að henni lokinni. Mælingar sem notaðar voru í rannsókninni voru málsýni og sérhannað mælitæki þar sem prófað var úr markorðaforðanum. Í töflu 1 má sjá yfirlit yfir þær mælingar sem gerðar voru.

Tafla 1. Yfirlit yfir mælingar meðan á rannsókninni stóð

Tími mælinga	Mælitæki
Mælingar áður en kennsla hefst	Tvö málsýni (samtalsmálsýni)
Íhlutunar/kennsluskeið: Daglegar mælingar	Sérhannað markorðapróf
Íhlutunar/kennsluskeið: Vikulegar mælingar	Málsýni (samtalsmálsýni)
Seinna grunnskeið: Mælingar strax að lokinni kennslu	Tvö málsýni (samtalsmálsýni)

Málsýni

Málsýni eru stutt sýnishorn af máltjáningu barna við eðlilegar aðstæður. Við úrvinnslu málsýna eru skoðaðir mælikvarðar á borð við meðallengd segða, fjölbreytileiki orða eða orðmynda, málfræði og setningagerð. Í þessari rannsókn var fylgt leiðbeiningum um töku málsýna í sjálfsprottnu tali og afritun þeirra sem er að finna í handbók um málsýni (Jóhanna Thelma Einarsdóttir og Þóra Sæunn Úlfsdóttir, 2018). Sjálfsprottið tal var laðað fram með því að spjalla við barnið í leik. Samræðurnar fylgdu áhuga barnsins. Stuðst var við opnar spurningar líkt og „Hvað gerðuð þið í útiveru í dag?“ fremur en að nota lokaðar spurningar á borð við „Var ekki gaman í útiveru í dag?“. Hvert málsýni var 50 segðir og voru niðurstöður bornar saman við viðmiðunartölur frá 39 börnum í aldurshópnum fimm og hálfs árs til fimm ára og ellefu mánaða ára (Jóhanna T. Einarsdóttir og Álfhildur Þorsteinsdóttir, 2015). Hugbúnaðurinn Málgreinir var notaður við úrvinnslu málsýnanna (Jóhanna Thelma Einarsdóttir og Stefán Carl Peiser, 2016). Í þessari grein eru skoðaðar niðurstöður mælikvarða fyrir fjölda mismunandi orðmynda og hlutfall málfræðivillna sem voru í flestum tilfellum beygingarvillur. Samtals voru tekin tíu málsýni hjá hvoru barni. Tvö málsýni voru tekin fyrir íhlutun, eitt málsýni vikulega á meðan á íhlutun stóð (alls sex málsýni) og tvö strax að lokinni íhlutun.

Sérhannað mælitæki

Hannað var sérstakt orðaforðapróf til að nota við rannsóknina að fyrirmynd prófs úr rannsókn Coynes o.fl. (2007). Prófað var í lok hvers dags úr markorðaforða dagsins á meðan á íhlutunarskeiði stóð, hvort sem um var að ræða beina eða óbeina kennslu. Þar sem markorðaforðinn var mismunandi eftir bókum var í raun um sex hliðstæð próf að ræða, sjá töflu 2. Börnin voru beðin um að skilgreina markorðin. Hægt var að fá tvö stig fyrir fullgilt svar, eitt stig fyrir sæmilegt svar en núll stig fyrir að snúa út úr eða svara ekki. Sem dæmi um fullgilt svar við markorðið *hrasa* var að *detta* en sem dæmi um sæmilegt svar var: „Bróðir minn dettur stundum“. Sérhannaða mælitækið var notað til þess að kanna þekkingu barnanna á markorðunum.

Val á sögubókum og markorðum

Í samráði við deildarstjóra barnanna í leikskólanum voru valdar bækur sem þóttu henta fimm ára börnum vel og sem ekki var verið að lesa fyrir börnin á meðan á rannsókninni stóð. Alls voru sex bækur lesnar fyrir börnin, valin voru sex markorð úr hverri bók eða alls 36 markorð, sjá nánar í töflu 2. Markorðin voru valin samkvæmt fyrirmynd Becks o.fl. (2013) úr öðru lagi orðaforðans (Tier-2). Í samráði við deildarstjóra á leikskóla barnanna og með hjálp íslenskrar orðabókar voru útskýringar búnar til á markorðunum sem hentuðu aldri og daglegri málnotkun barnanna. Óháður aðili sem hafði mikla reynslu í starfi með börnum á þessum aldri fór yfir útskýringarnar og val á markorðum. Valin voru tvö nafnorð, tvö sagnorð og tvö lýsingarorð úr hverri bók eins og gert var í rannsókn Coynes o.fl. (2007).

Tafla 2. Bækur, markorð og kennsla

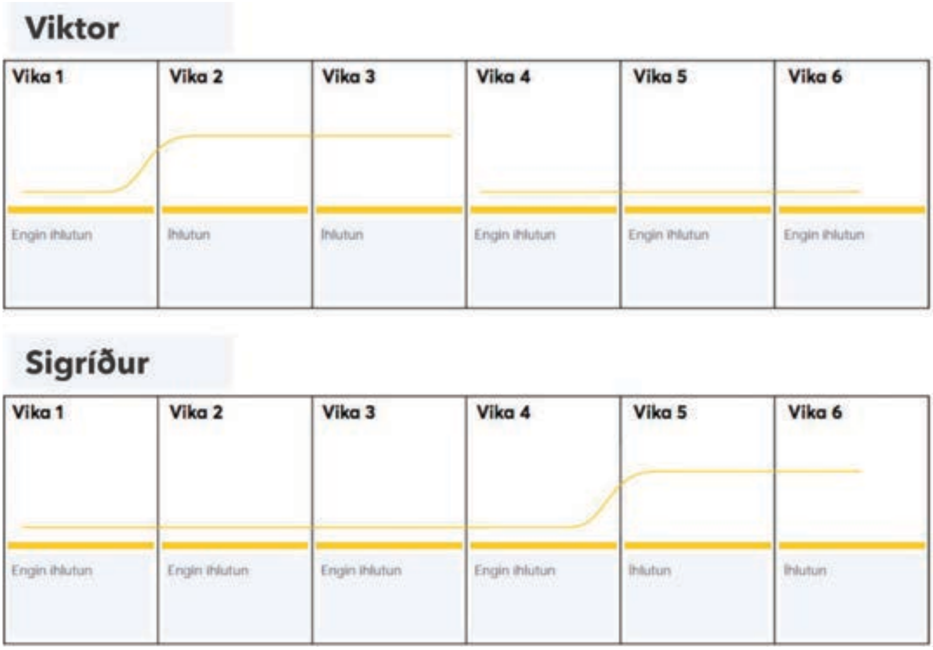
Heiti bókar	Höfundur/-ar	Markorð	Aðferð
Búkolla	Huginn Þór Grétarsson og Rosaria Battiloro	*ástúð, bál, baula, míga, stórstígur, ógnarstór*	*Mánudagur, vika 1, 2, og 3* Óbein kennsla Viktor (vika 1) Bein kennsla Viktor (vika 2 og 3) Óbein kennsla Sigríður (vika 1, 2 og 3)
Kuggur í sveitinni	Sigrún Eldjárn	*kunningi, öskuhaugar, æpa, aðstoða, undraverður, óþolinmóður.*	*Miðvikudagur, vika 1, 2, og 3* Óbein kennsla Viktor (vika 1) Bein kennsla Viktor (vika 2 og 3) Óbein kennsla Sigríður (vika 1, 2 og 3)
Amma fer í sumarfrí	Björk Bjarkardóttir	*áttaviti, vasaþjófur, hrópa, tylla, óvenjulega, troðfullur*	*Föstudagur, vika 1, 2, og 3* Óbein kennsla Viktor (vika 1) Bein kennsla Viktor (vika 2 og 3) Óbein kennsla Sigríður (vika 1, 2 og 3)
Höldum veislu, Einar Áskell!	Gunilla Bergström	*terta, nafnspjald, bragða, gægjast, himinlifandi, ljúffengur*	*Mánudagur, vika 4, 5, og 6* Óbein kennsla Sigríður (vika 1) Bein kennsla Sigríður (vika 5 og 6) Óbein kennsla Viktor (vika 1, 2 og 3)
Þegar ég verð stór ætla ég að spila með íslenska landsliðinu	Gemma Cary, Tatio Viana	*treyja, feðgar, hrósa, hrasa, splunkunýr, hjálpsamur*	*Miðvikudagur, vika 4, 5, og 6* Óbein kennsla Sigríður (vika 1) Bein kennsla Sigríður (vika 5 og 6) Óbein kennsla Viktor (vika 1, 2 og 3)
Hundurinn, kötturinn og músin	Bárður Óskarsson	*bæli, flótti, gabba, leiðast, grjótfúll, ógurlega*	*Föstudagur, vika 4, 5, og 6* Óbein kennsla Sigríður (vika 1) Bein kennsla Sigríður (vika 5 og 6) Óbein kennsla Viktor (vika 1, 2 og 3)

Framkvæmd í hnotskurn

Þjálfunaraðferðin sem notuð er í þessari rannsókn byggist á aðferð Coynes o.fl. (2007) um kennslu markorða með lestri sögubóka en með breyttu sniði. Þátttakendur voru færri, rannsóknartími lengri, fleiri bækur lesnar og fleiri markorð kennd. Bein kennsla fór þannig fram að áður en lestur hófst voru markorðin sex kynnt til leiks og barnið beðið um að láta vita ef það heyrði eitt af þessum orðum koma upp á meðan á lestri stæði. Ef barnið lét vita var setningin lesin aftur og útskýringin á markorðinu gefin. Ef barnið lét ekki vita var sú setning sem innihélt markorðið lesin aftur og barnið fékk beina útskýringu á því orði. Við óbeina kennslu voru bækurnar lesnar án útskýringa á orðunum. Markorðaforði dagsins var prófaður daglega með því að spyrja um skilgreiningu á markorðunum að loknum lestri, hvort sem um var að ræða beina kennslu eða óbeina. Kennslan fór fram í herbergi í leikskóla barnanna þar sem engin truflun var. Hver lestrarstund tók um 8–20 mínútur, óbeina kennslan tók styttri tíma en beina kennslan.

Rannsóknarsnið

Rannsóknarsniðið var einliðasnið með margliða grunnskeiði. Einliðasnið byggist á endurteknum mælingum þar sem kunnátta einstaklings áður en kennsla hefst er borin saman við kunnáttu hans á meðan á kennslu stendur og eftir að kennslu lýkur (Guðrún Árnadóttir og Þorlákur Karlsson, 2003). Hér var sérhannaða orðaforðaprófið notað til að bera saman áhrif beinnar kennslu við óbeina kennslu á orðaþekkingu barnanna. Skoðuð voru áhrif beinnar kennslu endurtekið með því að athuga orðanám barnanna á mismunandi orðum eftir vikudögum og bera saman á milli þátttakenda beina og óbeina kennslu. Þegar bein kennsla hófst hjá fyrsta þátttakanda var orðanámið borið saman við kennslu hjá þátttakandanum sem fékk óbeina kennslu. Tafla 2 sýnir nákvæmlega hvaða orð og bækur var unnið með í hverri viku og hvort barnið fékk beina kennslu eða óbeina eftir vikudögum og vikum. Eins og sést í töflu 2 var hver bók lesin þrisvar sinnum á sama vikudegi á viku fresti. Eftir þrjár vikur var skipt um bækur. Á mynd 1 er rannsóknarsniðið skýrt frekar og þar sést að fyrstu viku



Mynd 1. Margliða grunnskeið sem var notað í rannsókninni

kennslunnar fengu bæði börnin óbeina kennslu. Næstu tvær vikurnar hlaut Viktor beina kennslu á meðan Sigríður fékk áfram óbeina kennslu. Í fjórðu viku fengu bæði börnin aftur óbeina kennslu. Á fimmtu og sjöttu viku hlaut Sigríður beina kennslu en Viktor óbeina kennslu.

Meðferðarheldni

Meðferðarheldni (e. *fidelity*) er mæling á því hversu nákvæmlega áætlun meðferðar er fylgt (Íris Ösp Bergþórsdóttir o.fl., 2021; Dusenbury o.fl., 2003; Slaughter o.fl., 2015). Til að geta sagt til um árangur meðferðar eða borið saman ólíkar meðferðir er nauðsynlegt að geta mælt með nákvæmni hvernig meðferð er komið til skila. Án mælinga á meðferðarheldni væri auðveldlega hægt að draga ályktanir um árangur meðferða sem ekki reynast réttar (Íris Ösp Bergþórsdóttir og Ingham, 2016). Með þjálfun, handbókum og eftirliti er unnt að hafa betri stjórn á því hvernig meðferð er komið til skila (Slaughter o.fl., 2015).

Mat á meðferðarheldni

Til að meta meðferðarheldni var stuðst við fjóra mælikvarða. Þeir eru: *magn* (e. *dosage*), *fastheldni* (e. *adherence*), *gæði* (e. *quality*) og *svörun* (e. *responsiveness*). Hver þessara mælinga er mikilvæg til að útskýra og sýna á trúverðugan hátt hvað skilar árangri í rannsókninni (Dusenbury o.fl., 2003). Í þessari rannsókn var *magn* kennslunnar mælt með því að skrá niður hversu oft börnin mættu, hversu lengi hver kennslutími stóð, bæði fyrir beinu kennsluna og óbeinu kennsluna, og að lokum hversu margar vikur þjálfunin tók. Lagt var upp með að þátttakendur myndu mæta þrisvar í viku í sex vikur og að beina kennslan tæki 15 mínútur og óbeina kennslan 10 mínútur. *Fastheldni* er skilgreind sem mælikvarði á hvort íhlutunin var veitt af nákvæmni. Í rannsókninni var metið hvort rannsakandi útskýrði markorðin á viðunandi hátt með þriggja punkta kvarða. Einn stóð fyrir „ekki viðunandi“, tveir stóð fyrir „gerir stundum“ og þrír stóð fyrir „viðunandi“. *Gæði* segja til um hversu vel kennslunni er komið til skila. Það var gert með því að meta hversu vel rannsakandi kom lestrinum frá sér. Metið var hvort rannsakandi hafi lesið markorðin af innlifun eða af áhugaleysi. Sami þriggja punkta kvarðinn var notaður og í fastheldninni. Að lokum var metið hversu vel þátttakandi brást við kennslunni og hvort hann tók virkan þátt í henni. *Svörunin* var mæld á þriggja punkta kvarða þar sem einn var „hlustar ekki, hefur ekki ánægju af lestrinum“, tveir var fyrir „hlustar af hálfum huga“ og þrír var „hefur ánægju af lestrinum og hlustar af athygli“ (Dusenbury o.fl., 2003).

Allir þjálfunartímarnir voru teknir upp á myndband með snjallsíma (iPhone 8) og skoðaði rannsakandi öll myndböndin og mat meðferðarheldnina. Auk þess var óháður meistaranemi í talmeinafræði fenginn til þess að fara yfir þriðjung myndbandanna og meta meðferðarheldnina. Myndböndin sem hann mat voru valin af handahófi. Að lokum voru þessar mælingar bornar saman og samræmi milli matsmanna skoðað.

Tölfræðileg úrvinnsla

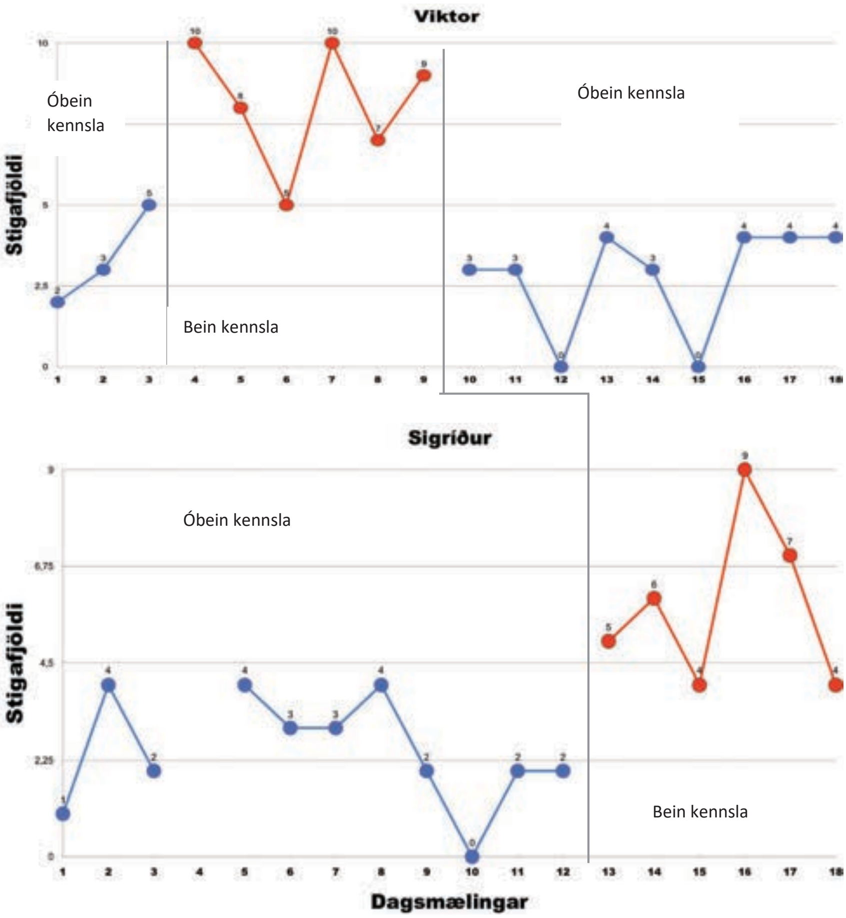
Gögnin voru skráð í Microsoft Word- og Excel-skjal. Til þess að skoða og meta gögnin var notuð sjónræn greining (e. *visual inspection*). Í þeirri greiningu eru mælingar bornar saman með línuriti og er hún gjarnan notuð þegar um einliðasnið er að ræða (Guðrún Árnadóttir og Þorlákur Karlsson, 2003).

Niðurstöður

Daglegar mælingar

Á mynd 2 má sjá niðurstöður úr sérhannaða orðaforðaprófinu sem lagt var fyrir daglega að loknum lestri. Borin er saman þekking barnanna á orðunum, annars vegar við beina kennslu og hins vegar við óbeina.

Mynd 2 sýnir daglegar mælingar á því hvað Sigríður og Viktor þekktu mörg af þeim sex markorðum sem voru kennd daglega. Mest var hægt að fá 12 stig eða tvö stig fyrir hvert markorð. Mismunandi bækur voru lesnar eftir vikudögum og um mismunandi markorð var að ræða eftir því hvaða sögu var verið að lesa, sjá töflu 2. Fyrri lotan (fyrstu þrjár vikurnar) hefst á óbeinni kennslu fyrstu vikuna (mælipunktar 1, 2, 3) og síðan fékk Viktor beina kennslu en Sigríður áfram óbeina kennslu (mælipunktar 4, 5, 6, 7, 8 og 9). Eins og sést á mynd 2 eru miklar framfarir í þekkingu á markorðunum hjá Viktori en ekki hjá Sigríði. Hjá Sigríði vantar mælingar frá fjórða degi vegna veikinda. Seinni lotan (seinni þrjár vikurnar) hefst á óbeinni kennslu (mælipunktar 10, 11 og 12) fyrir bæði börnin, síðan fær Sigríður beina kennslu í tvær vikur en Viktor óbeina (mælipunktar 13, 14, 15, 16, 17 og 18). Miklar framfarir urðu á þekkingu Sigríðar á markorðunum eins og kemur fram á mynd 2. Hliðstæðar fram-



Mynd 2. Niðurstöður mælinga á þekkingu Sigríðar og Viktors á markorðum í lok hvers dags. Blátt sýnir óbeina kennslu en rautt beina kennslu.

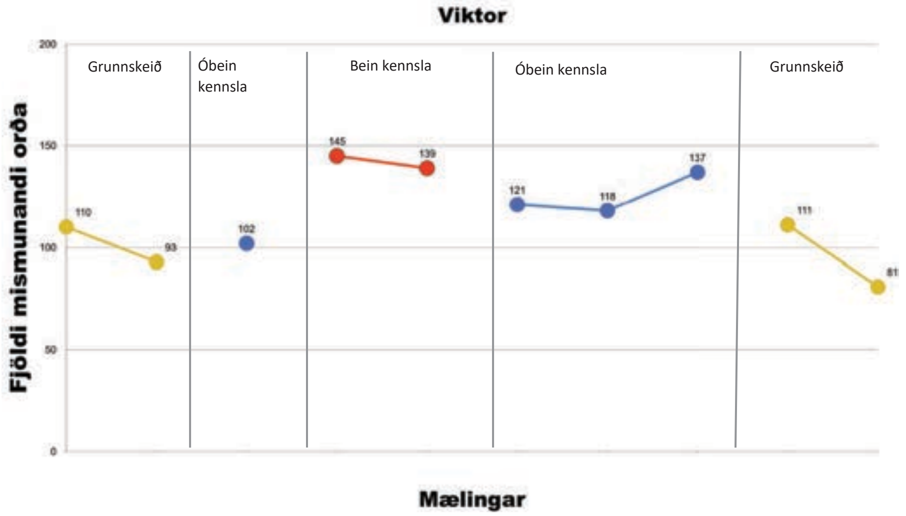
farir eru ekki hjá Viktori enda er um óbeina kennslu að ræða hjá honum og sagan lesin án útskýringa á markorðum. Niðurstöður úr dagsmælingum sýna greinilega að bein kennsla hafði skilað betri árangri en óbein kennsla. Einnig er hægt að draga þá ályktun að óbein kennsla hafi skilað litlum árangri, þ.e. hafi ekki bætt miklu við þekkingu barnanna á markorðunum.

Málsýni

Á grunnskeiði fyrir kennslu voru tekin tvö málsýni hjá hvoru barni. Meðan á kennslu stóð voru tekin vikuleg málsýni, eitt málsýni hjá hvoru barni í sex vikur. Að lokinni kennslu voru tekin tvö málsýni hjá hvoru barni. Samtals voru þetta tíu málsýni hjá hvoru barni. Öll málsýnin voru álíka löng eða 50–51 segð.

Fjöldi mismunandi orða

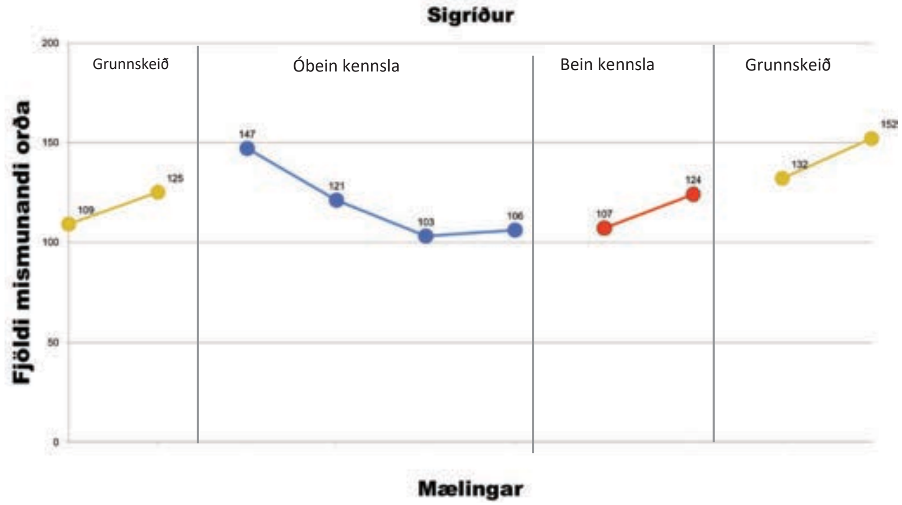
Á mynd 3 sést fjöldi mismunandi orða í málsýnum hjá Viktori en mynd 4 sýnir fjölda mismunandi orða í málsýnum hjá Sigríði.



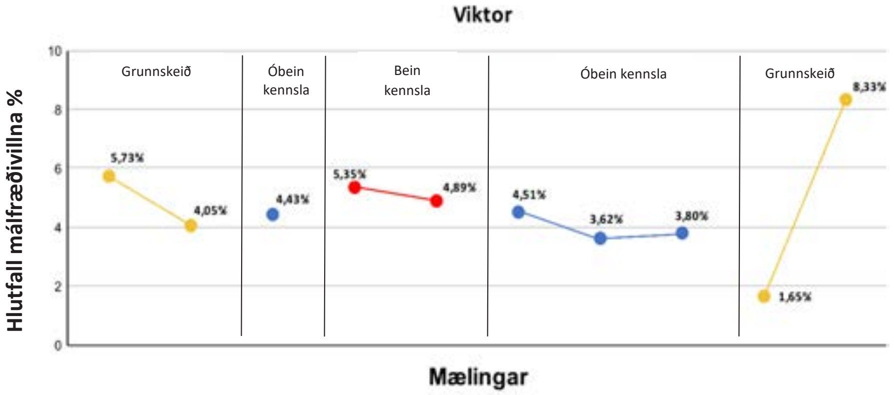
Mynd 3. Fjöldi mismunandi orða hjá Viktori. Gult sýnir grunnskeið, blátt stendur fyrir óbeina kennslu og rautt er bein kennsla

Nokkur dreifing er á fjölda mismunandi orða hjá Viktori eins og sést á mynd 4. Hann notaði frá 81 upp í 145 mismunandi orð í málsýnunum sem öll voru jafnlöng eða um 50 segðir. Sjá má að fjölbreytileiki orða í máltjáningu Viktors var mestur þær vikur sem hann fékk beina kennslu (139–145 orð). Myndin sýnir enn fremur að þó að bein kennsla hafi jákvæð áhrif á fjölda mismunandi orðmynda sem Viktor notar í sjálfsprottnu tali þá virðast áhrifin ekki vera varanleg því á seinna grunnskeiði eftir að kennslu lauk notar Viktor svipaðan fjölda mismunandi orða og áður en kennslan hófst.

Eins og sést á mynd 4 er fjöldi mismunandi orða hjá Sigríði nokkuð breytilegur eða frá 103 upp í 152 mismunandi orð, lægstur við óbeina kennslu en hæstur að lokinni kennslu. Kennslan virðist hafa haft jákvæð áhrif á fjölda mismunandi orða í sjálfsprottnu tali Sigríðar þannig að eftir kennsluna tjáir hún sig með fjölbreyttari orðum.



Mynd 4. Fjöldi mismunandi orða hjá Sigríði. Gult sýnir grunnskeið, blátt stendur fyrir óbeina kennslu og rautt er bein kennsla



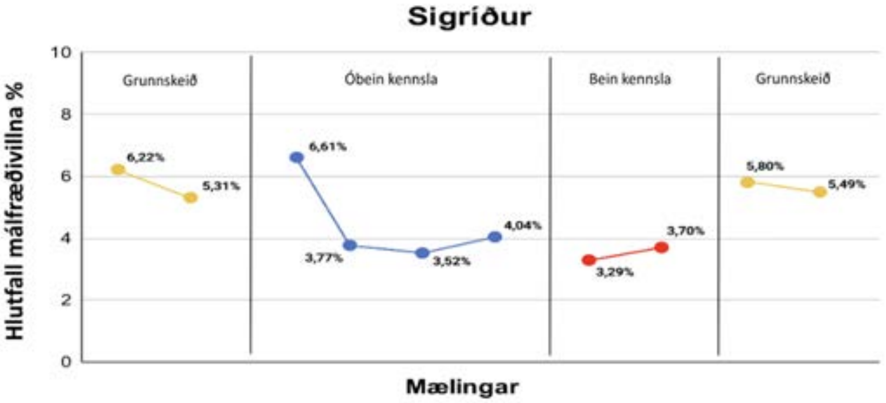
Mynd 5. Hlutfall málfræðivillna í málsýnum hjá Viktori. Gult sýnir grunnskeið, blátt stendur fyrir óbeina kennslu og rautt er bein kennsla

Hlutfall málfræðivillna

Á mynd 5 er hlutfall málfræðivillna í málsýnum hjá Viktori en á mynd 6 er hlutfall málfræðivillna í málsýnum hjá Sigríði.

Eins og sjá má á mynd 5 eru töluverðar sveiflur í hlutfallslegum fjölda villna hjá Viktori á seinna grunnskeiði eða frá 1,65% til 8,33%. Að meðaltali voru þær 4,6%, á grunnskeiði 4,94% og meðan á kennslu stóð 4,43% en miklar sveiflur voru í mælingum sérstaklega á seinna grunnskeiði.

Á mynd 6 má sjá nokkrar sveiflur í málfræðivillum hjá Sigríði í sjálfsprottnu



Mynd 6. Hlutfall málfræðivillna í málsýnum hjá Sigríði. Gult sýnir grunnskeið, blátt stendur fyrir óbeina kennslu og rautt er bein kennsla.

tali. Fyrstu vikuna á grunnskeiði var villufjöldinn hvað mestur eða um 6%. Næstu þrjár vikur við óbeina kennslu lækkar hlutfallslegur villufjöldi úr 6% í rúm 3%. Hlutfall málfræðivillna lækkaði enn frekar þegar hún hlaut beina kennslu og var þá lægstur um 3,3%. Að lokinni íhlutun á seinna grunnskeiði má sjá hækkun á hlutfalli málfræðivillna. Að meðaltali var hlutfall villna 4,8%, að meðaltali 5,7% á grunnskeiði en á meðan á kennslu stóð var það 4,1%.

Meðferðarheldni

Meðferðarheldnin var metin með sjálfsmati rannsakanda með því að fara yfir upptökurnar. Rannsakandi mat *magn* kennslunnar en hún fól í sér 18 kennslutíma á hvern þátttakanda, í alls sex vikur, 6 tíma fyrir beina kennslu og 12 tíma fyrir óbeina kennslu. Sigríður mætti í 17 tíma, alla tímana með beinu kennslunni en 5 tíma af 6 þegar óbeina kennslan átti sér stað. Viktor mætti í alla 18 tímana. Meðaltal þjálfunartíma þar sem bein kennsla var veitt var 16:18 mínútur á meðan meðaltal óbeinnar kennslu var 9:19 mínútur. Rannsakandi mat *fastheldni* og *gæði* sem viðunandi í öllum tilfellum Hvað varðar *svörun* taldi rannsakandi að börnin hefðu haft ánægju af lestrinum í 88,6% tilfella en í 11,4% tilfella hlustuðu þau með hálfum huga.

Samræmi á niðurstöðum matsmanna

Rannsakandi og óháður matsmaður mátu magn kennslunnar og voru alltaf sammála um fjölda skipta sem þátttakendur fengu kennslu og um lengd kennslunnar. Þegar kom að meðallengd kennslunnar var 91,8% samræmi milli matsmanna í beinu kennslunni en 88% samræmi í óbeinu kennslunni. Við mat á fastheldni og gæðum var 100% samræmi milli matsmanna. Samræmi matsmanna var 93% við mat á svörun þátttakenda.

Umræða

Meginmarkmið rannsóknarinnar var að athuga hvort auka megi orðaforða tvítyngdra barna á Íslandi með því að lesa fyrir þau sögubækur. Borin voru saman áhrif beinnar og óbeinnar kennslu á orðanám á 36 markorðum sem voru valin úr sex barnabókum. Helstu niðurstöður og um leið ávinningur þessarar rannsóknar var að bein kennsla með lestri sögubóka skilaði betri árangri en óbein kennsla. Við beina kennslu voru markorðin útskýrð á meðan á lestrinum stóð en við óbeina kennslu voru markorðin ekki útskýrð sérstaklega. Þessar niðurstöður eru í samræmi við aðrar sambærilegar rannsóknir (Beck o.fl., 2013; Coyne o.fl., 2004; Justice o.fl., 2005; Sigrún Alda Sigfúsdóttir o.fl., 2020) sem sýna að við sögulestur er mikilvægt að staldra við einstök orð og útskýra þau vel um leið og þau koma fyrir í sögunni. Það ýtir undir að börn læri ný orð en hvetur þau einnig til að vera virk á meðan á lestrinum stendur. Enn fremur var kannað hvort sögulesturinn og kennsla á orðum skilaði sér í fjölbreytilegri orðanotkun og færri málfræðivillum í almennri máltjáningu barnanna eins og hún endurspeglaðist í málsýnum þeirra. Niðurstöður sýndu að orðaforðakennslan virtist hafa haft frekar jákvæð áhrif á notkun á fjölbreytilegum orðum í almennri máltjáningu barnanna en ekki afgerandi áhrif á hlutfallslegan fjölda málfræðivillna. Þess ber að geta að fjölbreytileiki orða í almennri máltjáningu barnanna var svipaður og oft meiri en hjá eintyngdum jafnöldrum. Hins vegar voru málfræðivillur hjá báðum börnunum töluvert fleiri en hjá eintyngdum jafnöldrum þeirra (Jóhanna T. Einarsdóttir og Álfhildur Þorsteinsdóttir, 2015).

Markorðaforði

Fyrsta rannsóknarspurningin fjallaði um hvort Viktor og Sigríður myndu ná að auka við orðaforða sinn með því að fá beina útskýringu á markorðum. Til þess að meta markorðaforðann var notað sérhannað orðaforðapróf sem kannaði þekkingu barnanna á markorðunum. Orðaforðaprófið var byggt á fyrirmynd úr rannsókn Coynes o.fl. (2007). Rannsóknarsniðið var sett upp í tvær þriggja vikna lotur og hófst hvor lota á óbeinni kennslu. Þannig var hægt að bera saman þekkingu á orðunum áður en bein kennsla hófst og enn fremur að skoða áhrif beinnar kennslu. Í fyrri lotunni fékk Viktor beina kennslu en ekki Sigríður og í seinni lotunni fékk Sigríður beina kennslu en ekki Viktor. Með því var hægt að bera saman orðanám á milli barnanna við beina og óbeina kennslu. Í viku hverri var unnið með þrjár bækur. Til að mynda var *Búkolla* tekin fyrir á mánudögum í fyrri lotu og unnið með orðin úr henni þann dag. Á miðvikudögum voru tekin fyrir ný orð úr bókinni *Kuggur í sveitinni*. Daglegar mælingar úr markorðunum sýndu jákvæðari áhrif beinnar kennslu en óbeinnar. Báðir þátttakendur voru samvinnuþýðir og áhugasamir um að læra ný orð. Þau heyrðu markorðin og útskýringu þeirra oft á meðan á íhlutun stóð, sem mögulega leiddi til þess að þau voru farin að læra hvernig þau gætu útskýrt orðin. Ef þau voru ekki viss um hvað orðið þýddi sem spurt var um voru þau hvött til þess að giska á merkingu þess Þessar niðurstöður um áhrif beinnar kennslu á orðaforða eru í takt við sambærilegar erlendar rannsóknir. Sem dæmi má nefna að Biemiller og Boote (2006) og Coyne o.fl. (2007) hafa sýnt fram á að börn læra fleiri ný orð með beinni kennslu en með óbeinni, þar sem þau þurfa að giska á merkingu orða út frá samhengi. Svipuð rannsókn, þar sem þátttakendur voru með DLD/málþroskaröskun sýndi fram á að bein kennsla væri árangursríkari en óbein (Sigrún Alda Sigfúsdóttir o.fl., 2020).

Mælingar með málsýnum

Seinni rannsóknarspurningin fól í sér að skoða áhrif kennslunnar á sjálfsprottið tal hjá börnunum með málsýnum. Niðurstöður bentu til þess að kennslan, bæði bein og óbein, hafi haft jákvæð áhrif á hversu fjölbreytt orð þau notuðu í málsýnunum. Hjá Viktori var fjölbreytileiki orða mestur á meðan á beinu kennslunni stóð. Fjölbreytileiki orða á grunnskeiði fyrir og eftir kennslu var nánast sá sami þannig að kennslan virtist ekki hafa haft áhrif á fjölda mismunandi orða ef til lengri tíma er litið. Viktor notaði að meðaltali 99 mismunandi orð á báðum grunnskeiðum en að meðaltali 127 mismunandi orð á meðan á kennslu stóð. Hins vegar notaði Sigríður fjölbreyttari orð eftir að kennslu lauk, eða 142 orð að meðaltali, en á fyrsta grunnskeiði (117 að meðaltali) og mun fjölbreyttari orð en eintyngdir jafnaldrar. Til samanburðar má geta þess að eintyngdir jafnaldrar nota 119 orð að meðaltali en eins og getið var um í inngangi er dreifing mikil innan hvers aldurshóps og staðalfrávikið hátt (Jóhanna T. Einarsdóttir og Álfhildur Þorsteinsdóttir, 2015). Því má álykta að kennslan, bæði bein og óbein, hafi haft fremur jákvæð áhrif á að auka fjölbreytileika orða í almennri máltjáningu hjá báðum börnunum. Hlutfallslegur fjöldi málfræðivillna var hins vegar mun meiri bæði hjá Sigríði og Viktori en að meðaltali hjá eintyngdum jafnöldrum og breyttist lítið á meðan á rannsókninni stóð. Hlutfall málfræðivillna var að meðaltali 4,8% hjá Sigríði en 4,6% hjá Viktori. Eintyngdir jafnaldrar gera mun færri málfræðivillur eða að meðaltali 1,6% (Jóhanna T. Einarsdóttir og Álfhildur Þorsteinsdóttir, 2015). Í rannsóknum þar sem notuð hafa verið málsýni til að kanna málfærni barna með önnur heimamál en íslensku hafa niðurstöður verið hliðstæðar, þ.e. málfræðivillur hjá þeim hafa verið fleiri en hjá eintyngdum jafnöldrum (Hjördís Hafsteinsdóttir o.fl., 2022).

Elín Þöll Þórðardóttir (2011) sýndi fram á að magn og gæði samskipta á hvoru tungumálinu fyrir sig skipta öllu máli fyrir málþroska tvítyngdra barna. Báðir þátttakendur þessarar rannsóknar vörðu meirihluta af vökutíma sínum í leikskóla í íslensku málumhverfi. Mælingar á málþroska þeirra sýndu að þau höfðu náð góðri færni í íslensku fyrir íhlutun á öllum þáttum fyrir utan málfræðina. Börnunum fannst gaman að taka þátt í rannsókninni, læra ný orð og hlusta á sögurnar eins og kom fram við athugun á meðferðarheldni rannsóknarinnar. Að sögn foreldra var málþekking þeirra á pólsku svipuð og hjá pólskum jafnöldrum þeirra. Hér er því gott dæmi um vel heppnað og virkt tvítyngi.

Hins vegar má velta því fyrir sér hvort efla mætti færni þeirra í íslensku enn frekar með því að vinna að því að fækka málfræðivillum sem þau gera í sjálfsprottnu tali. Sýnt hefur verið fram á að áhrifarík leið við að hafa áhrif á málfræðiþekkingu barna í daglegu tali er að endurtaka leiðrétt (e. *recasting*) það sem börnin segja (Cleave o.fl., 2015). Með jákvæðri endurtekningu heyra börnin rétta setningamyndun og beygingar orða í daglegu tali. Ef allt starfsfólk leikskólans verður meðvitaðra um að leiðrétta á jákvæðan hátt málfar barnanna í daglegu tali er það hugsanlega nægilegt til að börnin nái betri tökum á íslenskri málfræði.

Ávinningur rannsóknarinnar

Mikilvægt er að finna árangursríkar aðferðir til þess að auka við orðaforða og samtímis að efla málfærni ungra barna. Aðferðirnar þurfa að vera einfaldar í framkvæmd og geta nýst foreldrum, kennurum og öðrum uppalendum. Þessi rannsókn var gerð að fyrirmynd erlendrar rannsóknar (Coyne o.fl., 2007) þar sem lesið var fyrir hóp barna og voru niðurstöður hliðstæðar. Eins og hjá Coyne og félögum (2007) jókst orðaforði barnanna meira við beina kennslu en óbeina. Ekkert bendir til annars en að aðferðina væri hægt að nýta með hópi barna í almennu leikskólastarfi. Það mætti útfæra hana á mismunandi vegu með því að virkja barnahópinn í að velja orð og efla þau í því að ræða við og fræða hvert annað um orð og merkingu þeirra. Sýnt hefur verið fram á að það að lesa fyrir börn hefur jákvæð áhrif á málfærni þeirra. Að lesa og útskýra samhliða ákveðin markorð hefur jákvæðari áhrif á orðaforða barna en þegar þau reyna að geta sér til um merkingu orða út frá samhengi.

Gæði málörvunar í leikskólanum skiptir sköpum fyrir tvítyngd börn eigi þau að ná nægilega góðum tökum á skólamálinu, íslenskunni. Orðaforði er sérstaklega mikilvægur og er háður því hversu ríkuleg og innihaldsrík málörvunin er í umhverfi barnsins (Hart og Risley, 1995; Hoff o.fl., 2012; Noble o.fl., 2005). Sterk tengsl eru á milli stærðar orðaforða tvítyngdra barna og lesskilnings þeirra, sem er mikilvægur fyrir allt nám (Sigríður Ólafsdóttir o.fl., 2016). Börnin í þessari rannsókn töluðu bæði pólsku heima en íslensku í leikskólanum. Þau höfðu náð góðum tökum á báðum tungumálum fyrir utan íslenska málfræði. Málörvun þeirra í leikskólanum á íslensku hefur eflaust verið margþætt og góður grunnur var lagður áður en þessi rannsókn fór fram. Með þessari rannsókn jókst orðaforði barnanna enn frekar og þau lærðu markorðin sem voru útskýrð með beinni kennslu.

Takmarkanir rannsóknarinnar

Margir kostir fylgja því að nota einliðasnið og margliða grunnskeið þar sem hægt er að bera saman þekkingu einstaklinga fyrir og eftir íhlutun og einnig að bera saman á milli þátttakenda (Guðrún Árnadóttir og Þorlákur Karlsson, 2003). En hafa þarf í huga þætti sem gætu ógnað innra réttmæti rannsóknarinnar. Í þessari rannsókn var reynt að koma í veg fyrir mögulega áhrifaþætti eins og að börnin hefðu heyrt sögurnar og markorðin áður. Hugsanlega hefur það haft áhrif að sérhannaða mælitækið eða orðaforðaprófið var ekki forprófað áður en rannsóknin hófst frekar en í rannsókn Coynes o.fl. (2007). Í raun var um sex hliðstæð orðaforðapróf að ræða, eitt úr hverri bók. Í þessari rannsókn voru börnin tekin út úr barnahópnum og lesið fyrir þau einstaklingslega.

Hugsanlega yrðu niðurstöður aðrar ef lesið væri fyrir hóp barna og væri áhugavert að kanna það frekar.

Ályktanir

Hægt er að álykta að orðaforðakennsla með beinni aðferð, þar sem markorð eru útskýrð um leið og þau koma fyrir í sögubók sem lesin er, skili betri árangri en óbein kennsla þar sem markorð eru ekki útskýrð. Áhugavert rannsóknarefni væri að kanna hvernig almennt er lesið fyrir börn í leikskólum landsins og hvort staldrað sé við og orðin útskýrð sérstaklega. Einnig væri áhugavert að kanna hvort og hvernig markorð eru valin sem kynnt eru fyrir börnunum áður en lesturinn hefst. Í þessari rannsókn voru sex orð valin úr hverri bók eða 18 orð á viku. Börnin réðu vel við að læra svo mörg ný orð í viku hverri að viðhafðri beinni kennslu, nokkuð sem mætti taka til greina þegar verið er að lesa fyrir börnin. Samræðulestur með beinni kennslu er ein leið til að efla málfærni barna og dýpka þekkingu þeirra á tungumálinu. Í barnabókum og rituðum texta koma fyrir flóknari orð og setningar en í daglegu talmáli. Því er brýnt að nota barnabækur til að efla og auka orðaforða barnanna. Sérstaklega er mikilvægt að kynna ný orð fyrir börnum með annað móðurmál en íslensku og útskýra þau sérstaklega þar sem þau hafa yfirleitt minni orðaforða í íslensku en eintyngd börn (Aneta Figlarska o.fl., 2017; Hjördís Hafsteinsdóttir o.fl., 2021). Til að efla málfærni allra barna er góð málörvun í almennu starfi leikskólans lykilatriði. Í leikskólanum er lagður grunnur að lestri og lesskilningi og þar með öflugri málfærni á fullorðinsárum.

Heimildir

Amalía Björnsdóttir, Jóhanna T. Einarsdóttir og Ingibjörg Símonardóttir. (2013). Langtímarannsókn á forspárgildi málþroskaathugana við 5–6 ára aldur um síðari líðan og reynslu. *Netla – Veftímarit um uppeldi og menntun.* http://netla.hi.is/greinar/2013/ryn/013.pdf

Aneta Figlarska, Rannveig Oddsdóttir, Samúel Lefever og Hrafnhildur Ragnarsdóttir. (2017). Pólskur og íslenskur orðaforði tvítyngdra leikskólabarna: Málumhverfi heima og í leikskóla. *Netla – Veftímarit um uppeldi og menntun.* http://netla.hi.is/serrit/2017/menntakvika_2017/007.pdf

Árdís Hrönn Jónsdóttir. (2013). *Orðaspjall: Að efla orðaforða og hlustunarskilning barna með bókalestri.* Leikskólinn Tjarnarsel.

Beck, I. L. og McKeown, M. G. (2007). Increasing young low-income children's oral vocabulary repertoires through rich and focused instruction. *Elementary School Journal, 107*(3), 251–271. https://doi.org/10.1086/511706

Beck, I. L., McKeown, M. G. og Kucan, L. (2008). *Creating robust vocabulary: Frequently asked questions and extended examples.* Guilford Press.

Beck, I. L., McKeown, M. G. og Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction* (2. útgáfa). Guilford Press.

Beck, I. L., McKeown, M. G. og Omanson, R. C. (1987). The effects and uses of diverse vocabulary instructional techniques. Í M. G. McKeown og M. E. Curtis (ritstjórar), *The nature of vocabulary acquisition* (bls. 147–163). Lawrence Erlbaum.

Beck, I. L., Perfetti, C. A. og McKeown, M. G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology, 74*(4), 506–521. https://doi.org/10.1037/0022-0663.74.4.506

Biemiller, A. og Boote, C. (2006). An effective method for building meaning vocabulary in primary grades. *Journal of Educational Psychology, 98*(1), 44–62. https://doi.org/10.1037/0022-0663.98.1.44

Brooks, P. J. og Kempe, V. (2012). *Language development*. Wiley-Blackwell.

Cleave, P. L., Becker, S. D., Curran, M. K., Van Horne, A. J. O. og Fey, M. E. (2015). The efficacy of recasts in language intervention: A systematic review and meta-analysis. *American Journal of Speech-Language Pathology, 24*(2), 237–255. https://doi.org/10.1044/2015_AJSLP-14-0105

Coyne, M. D., McCoach, D. B. og Kapp, S. (2007). Vocabulary intervention for kindergarten students: Comparing extended instruction to embedded instruction and incidental exposure. *Learning Disability Quarterly, 30(*2), 74–88. https://doi.org/10.2307/30035543

Coyne, M. D., Simmons, D. C., Kame'enui, E. J. og Stoolmiller, M. (2004). Teaching vocabulary during shared storybook readings: An examination of differential effects. *Exceptionality, 12*(3), 145–162. https://doi.org/10.1207/s15327035ex1203_3

Dusenbury, L., Brannigan, R., Falco, M. og Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research, 18*(2), 237–256. https://doi.org/10.1093/her/18.2.237

Elín Þöll Þórðardóttir. (1998). *Orðaskil: Málþroskapróf: Leiðbeiningar og aldursviðmið.* Framsaga.

Elín Þöll Þórðardóttir. (2011). The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism, 15*(4), 426–445. https://doi.org/10.1177/1367006911403202

Elín Þöll Þórðardóttir. (2021). Adolescent language outcomes in a complex trilingual context: When typical does not mean unproblematic. *Journal of Communication Disorders,* 89, 106060.

Elín Þöll Þórðardóttir og Anna Guðrún Júlíusdóttir. (2013). Icelandic as a second language: A longitudinal study of language knowledge and processing by school-age children. *International Journal of Bilingual Education and Bilingualism, 16*(4), 411–435. https://doi.org/10.1080/13670050.2012.693062

Grøver, V., Rydland, V., Gustafsson, J. E. og Snow, C. E. (2020). Shared book reading in preschool supports bilingual children's second-language learning: A cluster-randomized trial. *Child Development, 91*(6), 2192–2210.

Guðrún Árnadóttir og Þorlákur Karlsson. (2003). Einliðasnið: Öflug leið til samhæfingar klínískrar vinnu og rannsókna. Í Sigríður Halldórsdóttir og Kristján Kristjánsson (ritstjórar), *Handbók í aðferðafræði og rannsóknum í heilbrigðisvísindum* (bls. 295–329). Háskólinn á Akureyri.

Hagstofa Íslands. (2022a). Börn í leikskólum eftir aldri, lengd viðveru og rekstrarform 1998–2020. https://px.hagstofa.is/pxis/pxweb/is/Samfelag/Samfelag__skolamal__1_leikskolastig__0_lsNemendur/SKO01001.px

Hagstofa Íslands. (2022b). Börn í leikskólum með erlent móðurmál 1998–2020. https://px.hagstofa.is/pxis/pxweb/is/Samfelag/Samfelag__skolamal__1_leikskolastig__0_lsNemendur/SKO01103.px

Hart, B. og Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Paul H. Brookes.

Hjördís Hafsteinsdóttir, Jóhanna Thelma Einarsdóttir og Íris Edda Nowenstein. (2022). Íslenskukunnátta tvítyngdra barna: Tengsl staðlaðra málþroskaprófa og málsýna. *Netla – Veftímarit um uppeldi og menntun.*

Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M. og Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language, 39*(1), 1–27. https://doi.org/10.1017/S0305000910000759

Hrafnhildur Ragnarsdóttir. (2015). Málþroski leikskólabarna: Þróun orðaforða, málfræði og hlustunarskilnings milli fjögra og fimm ára aldurs. *Netla – Veftímarit um uppeldi og menntun.* http://netla.hi.is/greinar/2015/ryn/007.pdf

Ingibjörg Símonardóttir, Einar Guðmundsson, Sigurgrímur Skúlason og Sigríður Pétursdóttir. (1995). *TOLD-2P: Málþroskapróf, handbók, íslensk staðfærsla.* Rannsóknastofnun uppeldis- og menntamála.

Íris Ösp Bergþórsdóttir og Ingham, R. J. (2016). Putting the cart before the horse: A cost effectiveness analysis of treatments for stuttering in young children requires evidence that the treatments analyzed were effective. *Journal of Communication Disorders, 65*, 65–67. https://doi.org/10.1016/j.jcomdis.2016.04.006

Íris Ösp Bergþórsdóttir, Kathryn Crowe og Jóhanna Thelma Einarsdóttir. (2021). Implementation fidelity in parent-implemented interventions for stuttering. *Clinical Linguistics & Phonetics, 36*, 1–24. https://doi.org/10.1080/02699206.2021.1965659

Jóhanna T. Einarsdóttir og Álfhildur Þorsteinsdóttir. (2015). Málsýni leikskólabarna: Aldursbundin viðmið. *Netla – Veftímarit um uppeldi og menntun*. http://netla.

hi.is/greinar/2015/ryn/010.pdf

Jóhanna T. Einarsdóttir, Amalía Björnsdóttir og Ingibjörg Símonardóttir. (2016). The predictive value of preschool language assessments on academic achievement: A 10-year longitudinal study of Icelandic children. *American Journal of Speech-Language Pathology, 25*(1), 67–79. https://doi.org/10.1044/2015_AJSLP-14-0184

Jóhanna Thelma Einarsdóttir og Stefán Carl Peiser. (2016). *Málgreinir.* http://malsyni.hi.is/

Jóhanna Thelma Einarsdóttir og Þóra Sæunn Úlfsdóttir. (2018). *Handbók um málsýni íslenskra barna.* http://malsyni.hi.is/

Justice, L. M., Meier, J. og Walpole, S. (2005). Learning new words from storybooks: An efficacy study with at-risk kindergartners. *Language, Speech, and Hearing Services in Schools, 36*(1), 17–32. https://doi.org/10.1044/0161-1461(2005/003)

Kriselle Lou Suson Jónsdóttir, Sigríður Ólafsdóttir og Jóhanna Thelma Einarsdóttir. (2018). Linguistically diverse children in Iceland: Their family language policy and Icelandic phonological awareness. Í Hanna Ragnarsdóttir og Samuel Lefever (ritstjórar), *Icelandic studies on diversity and social justice in education* (bls 7–38). Cambridge Scholars.

Marulis, L. M. og Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research, 80*(3), 300–335. https://doi.org/10.3102/0034654310377087

McKeown, M. G., Beck, I. L., Omanson, R. C. og Pople, M. T. (1985). Some effects of the nature and frequency of vocabulary instruction on the knowledge and use of words. *Reading Research Quarterly, 20*(5), 522–535. https://doi.org/10.2307/747940

Noble, K. G., Norman, M. F. og Farah, M. J. (2005). Neurocognitive correlates of socioeconomic status in kindergarten children. *Developmental Science, 8*(1), 74–87. https://doi.org/10.1111/j.1467-7687.2005.00394.x

Paradis, J. og Jia, R. (2016). Bilingual children's long-term outcomes in English as a second language: Language environment factors shape individual differences in catching up with monolinguals. *Developmental Science, 20*(1). e12433. https://doi.org/10.1111/desc.12433

Pearson, P. D., Hiebert, E. H. og Kamil, M. L. (2007). Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly, 42*(2), 282–296. https://doi.org/10.1598/RRQ.42.2.4

Rice, M. L., Buhr, J. C. og Nemeth, M. (1990). Fast mapping word-learning abilities of language-delayed preschoolers. *Journal of Speech and Hearing Disorders, 55*(1), 33–42. https://doi.org/10.1044/jshd.5501.33

Sigríður Ólafsdóttir. (2015). The development of vocabulary and reading comprehension among Icelandic second language learners [doktorsritgerð]. Skemman. http://hdl.handle.net/1946/23061

Sigríður Ólafsdóttir, Freyja Birgisdóttir, Hrafnhildur Ragnarsdóttir og Sigurgrímur Skúlason. (2016). Orðaforði og lesskilningur hjá börnum með íslensku sem annað mál: Áhrif aldurs við komuna til Íslands. *Netla – Veftímarit um uppeldi og menntun*. https://ojs.hi.is/netla/article/view/2428/1313

Sigríður Ólafsdóttir og Hrafnhildur Ragnarsdóttir. (2010). Íslenskur orðaforði íslenskra grunnskólanema sem eiga annað móðurmál en íslensku. *Ráðstefnurit Netlu – Menntakvika 2010*. https://netla.hi.is/serrit/2010/menntakvika2010/alm/023.pdf

Sigrún Alda Sigfúsdóttir, Jóhanna T. Einarsdóttir, Þorlákur Karlsson og Íris Ösp Bergþórsdóttir. (2020). Orðaforðakennsla með sögulestri fyrir börn með málþroskaröskun. *Netla – Veftímarit um uppeldi og menntun*. https://doi.org/10.24270/netla.2020.4

Slaughter, S. E., Hill, J. N. og Snelgrove-Clarke, E. (2015). What is the extent and quality of documentation and reporting of fidelity to implementation strategies: A scoping review. *Implementation Science, 10*(1), 129. https://doi.org/10.1186/s13012-015-0320-3

Spencer, E. J., Goldstein, H. og Kaminski, R. (2012). Teaching vocabulary in storybooks: Embedding explicit vocabulary instruction for young children. *Young Exceptional Children*, *15*(1), 18–32. https://doi.org/10.1177/1096250611435367

Stahl, S. A. og Nagy, W. E. (2006). *Teaching word meanings*. Lawrence Erlbaum.

Zevenbergen, A. A. og Whitehurst, G. J. (2003). Dialogic reading: A shared picture book reading intervention for preschoolers. Í A. Van Kleeck, S. A. Stahl og E. B. Bauer (ritstjórar), *On reading books to children: Parents and teachers* (bls. 170–195). Lawerence Erlbaum.

ÁGRIP

Rannsóknir benda til þess að íslenskur orðaforði tvítyngdra barna í leikskólum á Íslandi sé slakari en jafnaldra þeirra sem hafa íslensku að móðurmáli. Markmið rannsóknarinnar var að athuga orðanám barna með því að lesa sögu fyrir þau. Borin voru saman áhrif beinnar kennslu og óbeinnar. Þátttakendur voru tvö tvítyngd börn sem höfðu pólsku að móðurmáli og voru bæði á síðasta ári í leikskóla. Meðan á beinni kennslu stóð var staldrað við markorðið þegar það kom fyrir í textanum og það útskýrt. Við óbeina kennslu fengu börnin hefðbundinn sögulestur án útskýringar á markorðunum. Notað var einliðasnið með tveimur þriggja vikna lotum og kennt þrisvar sinnum í hverri viku. Niðurstöður leiddu í ljós að bein kennsla skilaði betri árangri en óbein kennsla. Enn fremur hafði kennslan jákvæð áhrif á fjölbreytileika orða í sjálfsprottnu tali barnanna. Niðurstöður rannsóknarinnar geta nýst starfsfólki leikskóla sem ein leið til að auka íslenskan orðaforða tvítyngdra barna.

ABSTRACT

Bilingual children, vocabulary and storybook reading

Previous research has indicated that bilingual children in Icelandic preschools have a smaller vocabulary in Icelandic than monolingual peers who are first-language learners of Icelandic. The aim of this study was to investigate children's vocabulary learning using a storybook-reading intervention. The impact of explicit teaching and indirect teaching approaches was compared regarding target vocabulary. The participants were two bilingual children who had Polish as their home language. Both children were in their last year of preschool (5–6 years old) in Iceland. In the direct teaching approach, each target word was explained when it appeared in the text. In the indirect approach, the story was read without explanation of the target words. The intervention used a single-case multiple baseline design, with two intervention blocks of three weeks duration each. Analysis of children's vocabulary knowledge showed that direct instruction was more effective than indirect instruction, and that the intervention (direct and indirect) had a positive effect on the number of different words the children used in spontaneous speech samples. This intervention approach may therefore be useful as an approach to increasing the Icelandic vocabulary of bilingual children.

Appendix C

Coding sheet for FOI measurements in the systematic review

Key information

- Author(s)
- Publication year
- Sample size
- participant age
- Name/type of methodology

Design/Method:

- Is there a clear statement of the aim of the research?
- Is It clear how the treatment will be organized and delivered?
- Can the study be replicated based on the design/description?

Training:

- Is the clinician's training mentioned and/or listed?
- Is there some form of supervision reported to prevent drift?
- Are the parents training described?

Dosage - Clinician

Reports of the actual number of sessions attended by the parents-child as well as the length of each session or a mean of session length. Documentation of when the intervention starts and finishes.

- Length of intervention (duration in weeks) (e.g. March 4th to September 8th)
- Number of intervention sessions (e.g. 35 sessions)
- length of sessions (e.g. 45 minutes on average each session)

Dosage - parent

Reports of the actual number of home-sessions as well as the length of each session.

- Length of intervention (duration in weeks) (e.g. March 4th to September 8th)
- Number of intervention sessions (e.g. 35 sessions)
- length of sessions (e.g. 45 minutes on average each session)

Adherence - Clinician

- Reporting of the use of any checklist, second observer, agreement data or other to track accuracy of implementation - was the intended intervention provided.
    - Self-administered checklist, observation, audio recording, video recording

Adherence - Parent

- Reporting of any measurement that indicates measures to determine that the intervention was provided in the home-sessions.

    - Audio recordings, video recordings, checklists, diary's, self-reports.

Quality - Clinician

- Reporting regarding how well the Intervention strategies were delivered at the clinicians sessions
    - Check-lists, second observer, manual, audio recording, video recording

Quality - Parent

- Reporting regarding how well the intervention strategies were delivered at the parent sessions
    - Check-lists, second observer, manual, audio recording, video recording

Responsiveness - Clinician

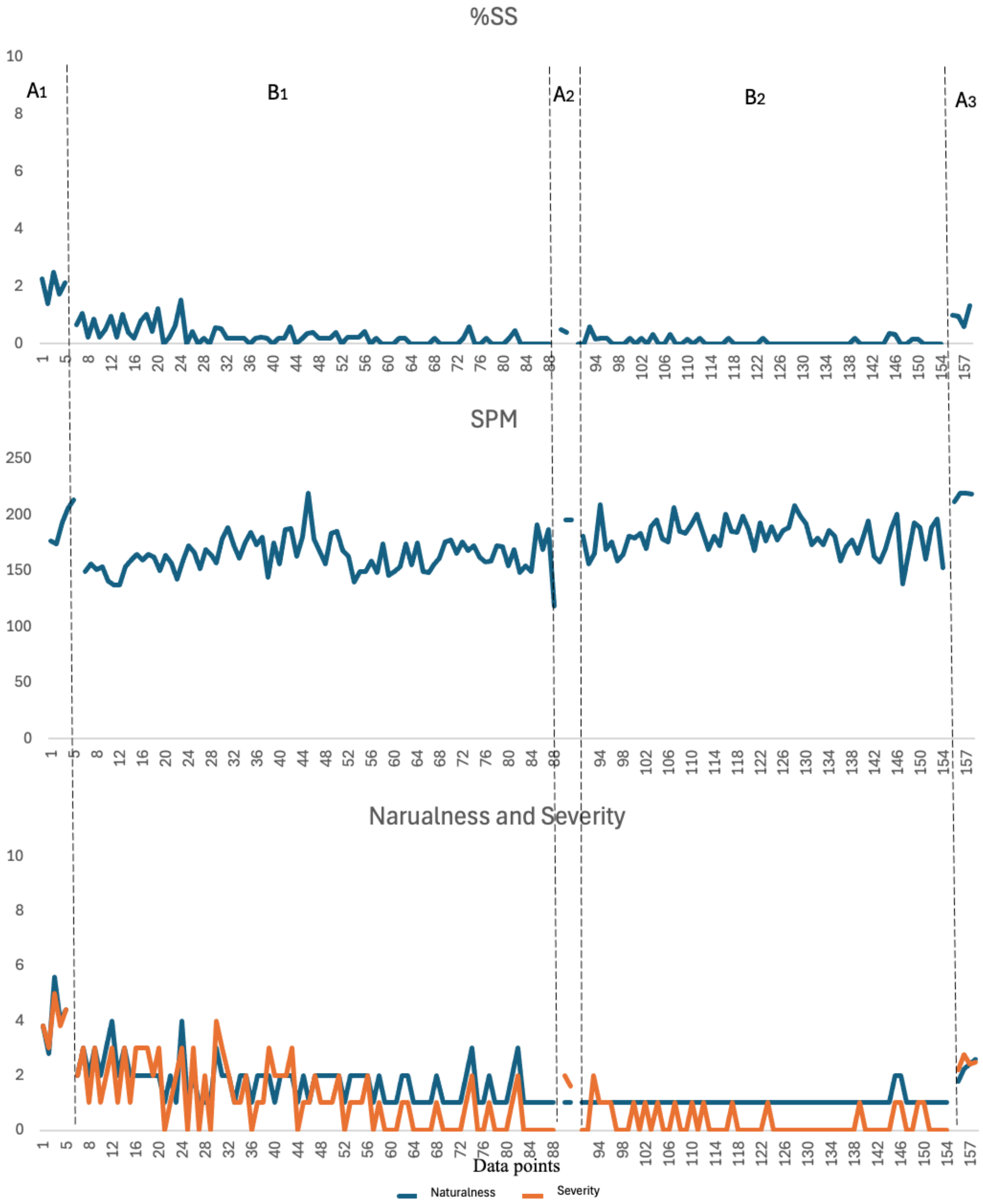
- Reports indicating to what extend parents are engaged and involved in the treatment
    - Information regarding attendance to treatment, parent satisfaction questionnaire, interviews

Responsiveness - Parent

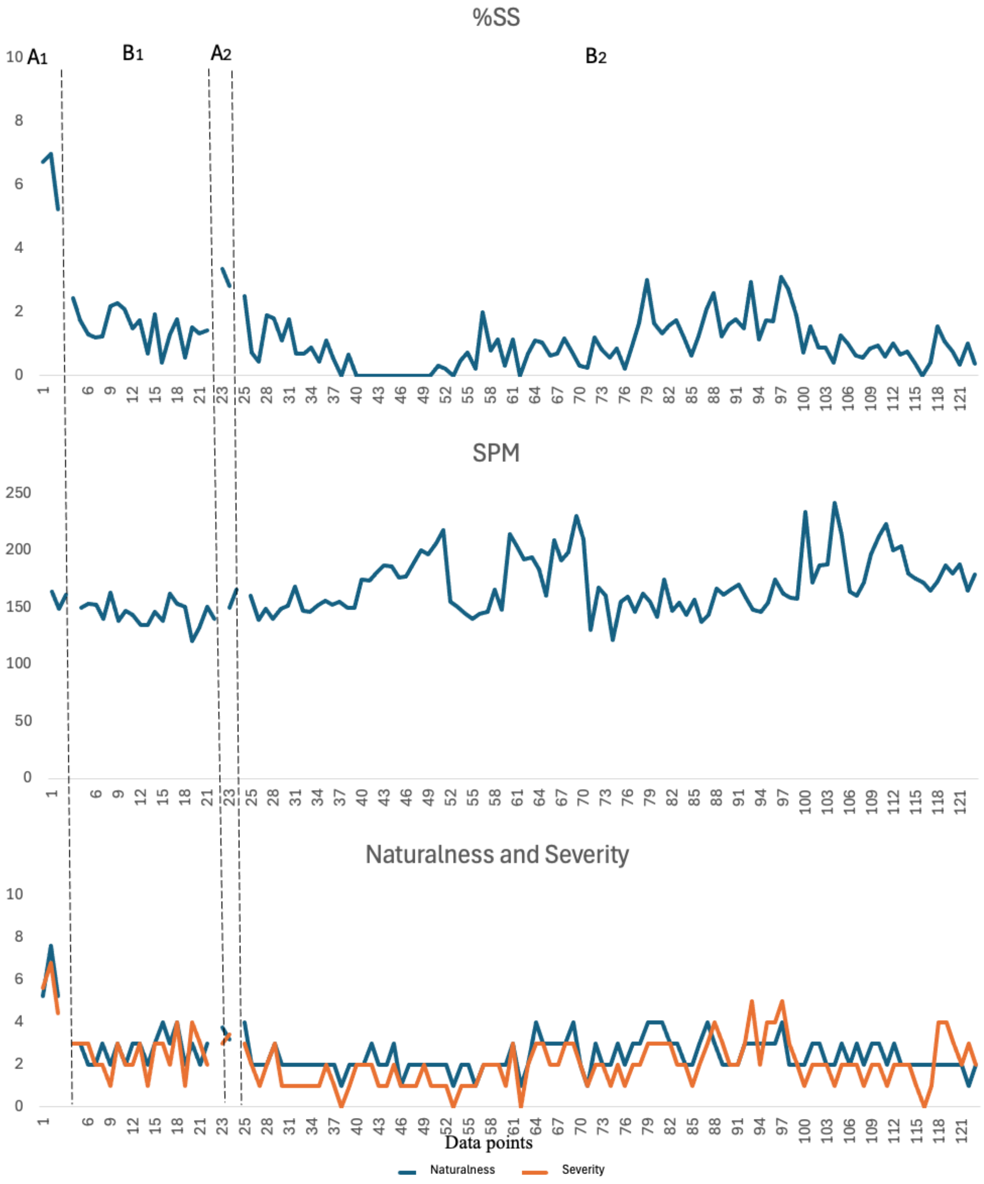
- Studies reported child behaviour during interactions with parent during treatment sessions or other observational measurements

Appendix D

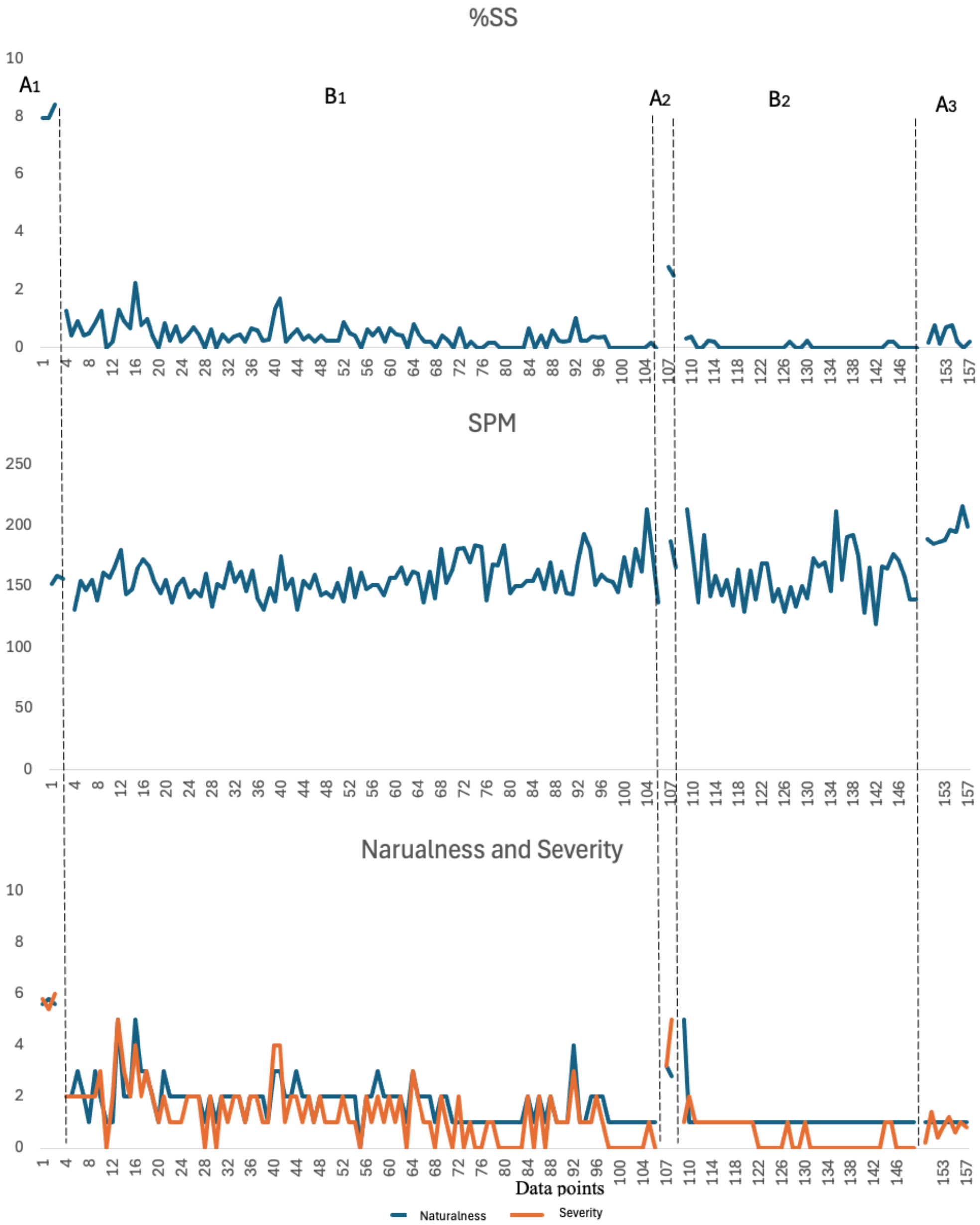
Monthly measurements (A) and intervention measurements (B) for P1



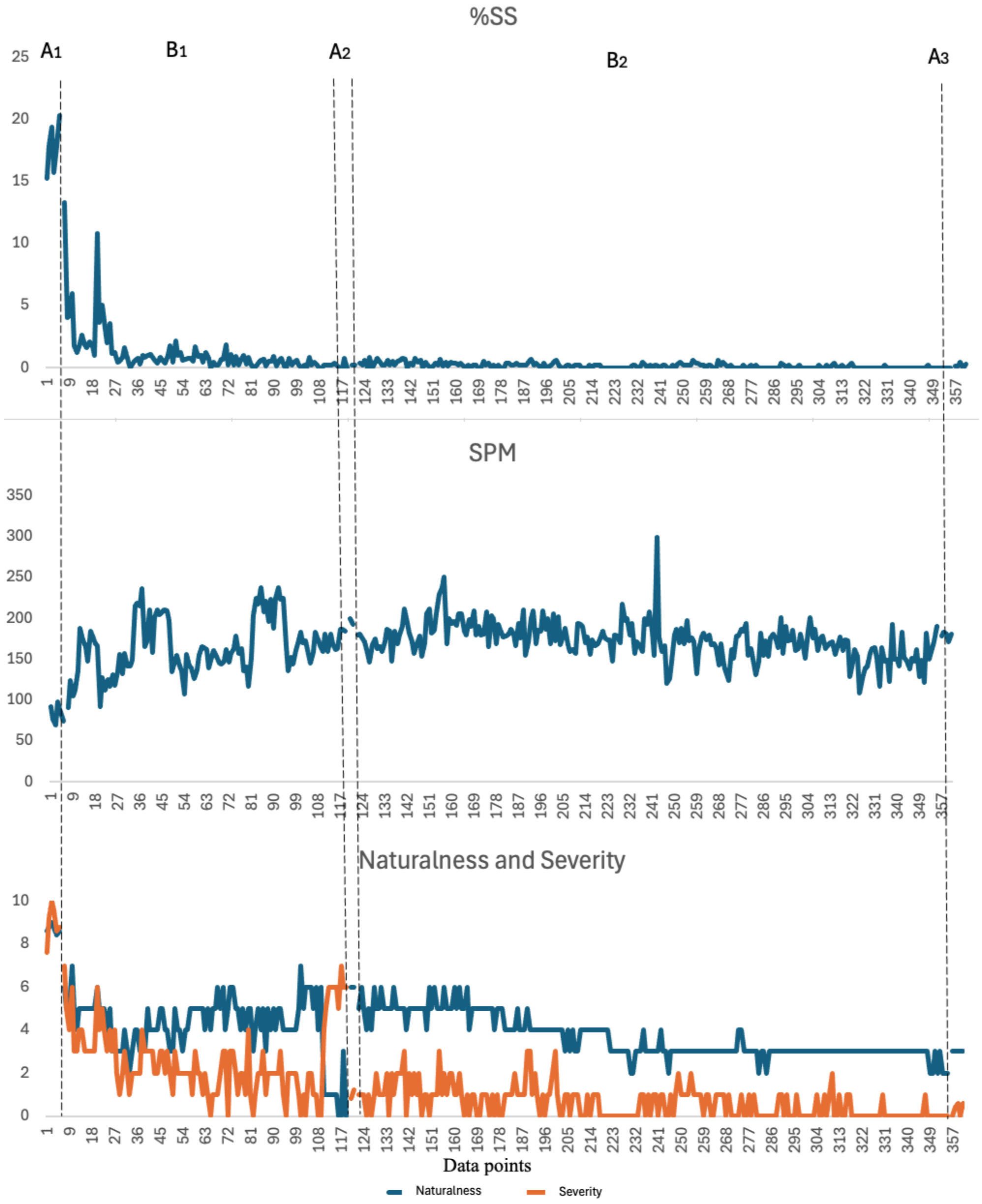
Monthly measurements (A) and intervention measurements (B) for P2



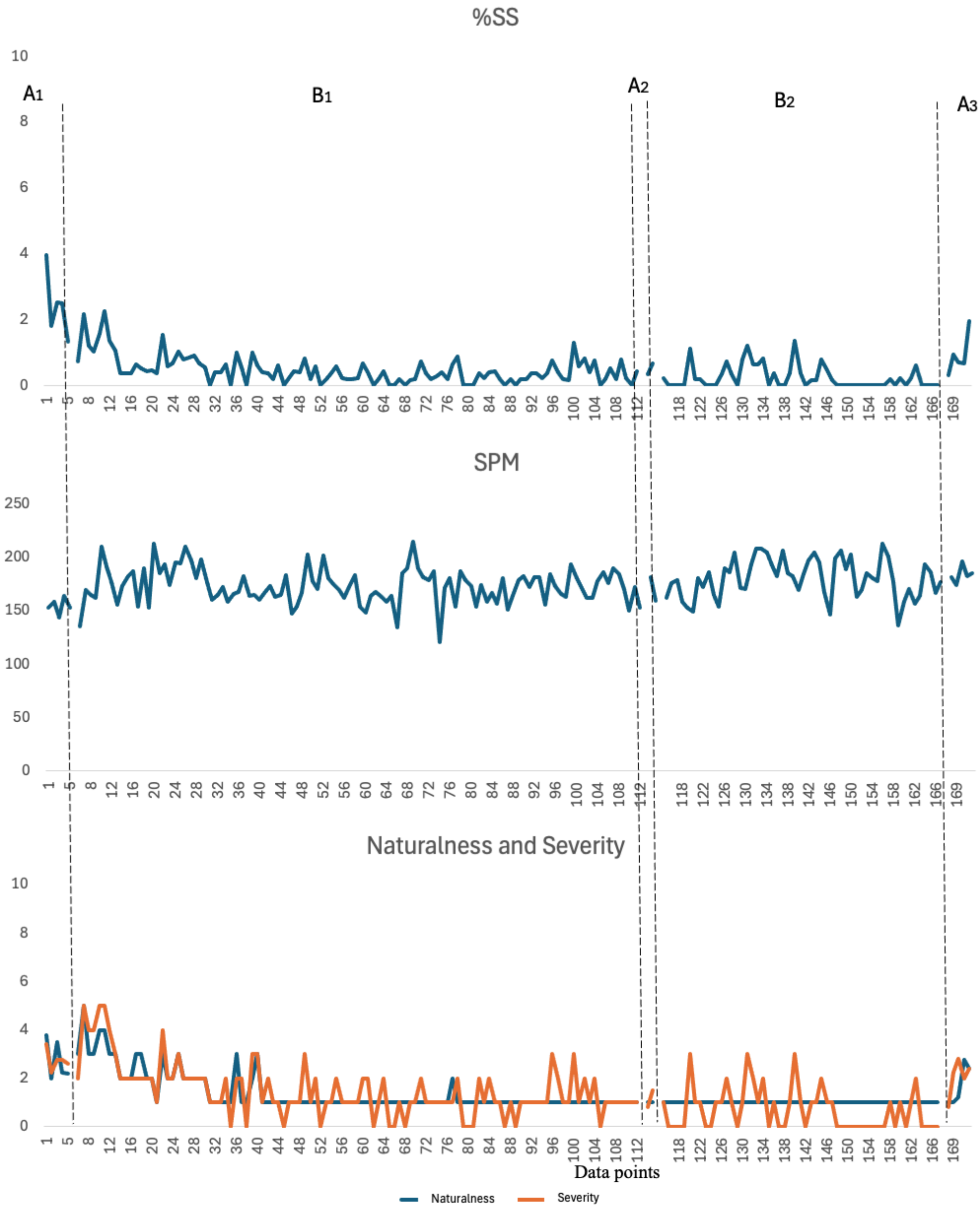
Monthly measurements (A) and intervention measurements (B) for P3



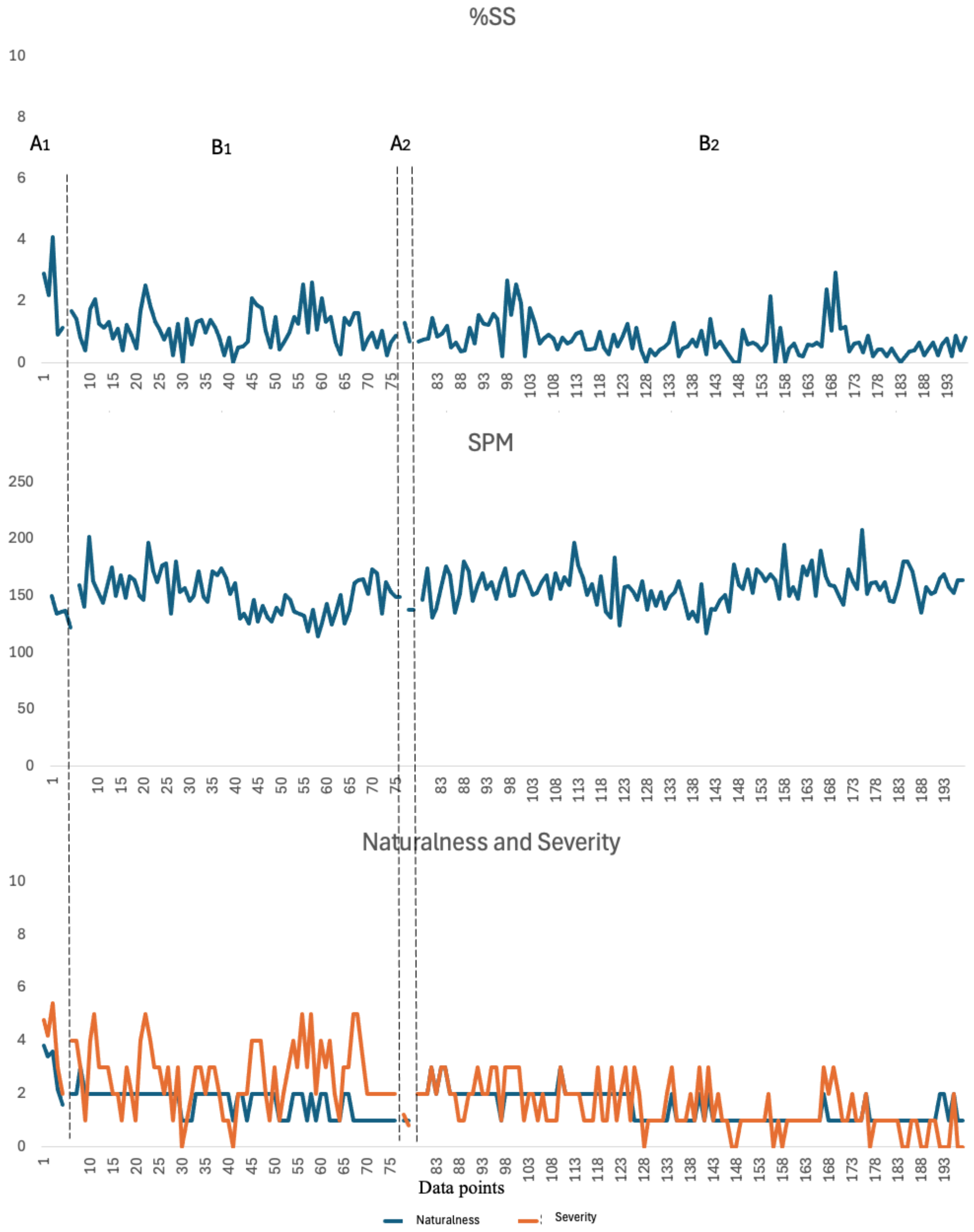
Monthly measurements (A) and intervention measurements (B) for P4



Monthly measurements (A) and intervention measurements (B) for P5



Monthly measurements (A) and intervention measurements (B) for P6



Monthly measurements (A) and intervention measurements (B) for P7