Supporting Second-Language Learning in the Early Elementary Years: A Researcher–Practitioner Approach

Mia Cecilie Heller

Thesis submitted for the degree of Ph.D.
Department of Education
Faculty of Educational Sciences

UNIVERSITY OF OSLO

2020

Supporting Second-Language Learning in the Early Elementary Years: A Researcher–Practitioner Approach

IV

Summary

Language skills provide essential building blocks for learning. Therefore, the lack of adequate language skills may have serious consequences in terms of schooling. In today's multicultural society, children enter school with varying degrees of language proficiency. Adapting instruction to variable levels of language skills is challenging. Thus, it is important to identify ways to foster language learning efficiently, specifically instructional support for language-minority learners who receive instruction in their second language. By focusing on explicit language instruction, teachers may enhance student's learning and prevent later literacy failure. Avenues for supporting language learning were examined in various ways through a researcher-practitioner partnership approach in the following three studies:

In paper 1, a randomized controlled trial (RCT) with language-minority learners in the early elementary years was conducted. The study aimed to support students' second-language (L2) skills by introducing them to an eight-week intervention program consisting of 64 lessons. The main components comprised semantic categories and their semantically related words, basic sentence production, and invitations to extended talk, all of which were supported with visual material. The sample consisted of 137 students (mean age = 6.3 years) in first and second grade. Children were randomly allocated to either an intervention group or a waiting-list control group. Students' language skills were assessed at three time points: pre-test, post-test, and follow up four months later. The results revealed that the intervention group showed an overall effect size of 0.35 compared to the waiting-list control group. After the waiting-list control group had received the intervention, no significant differences were found between the groups.

In paper 2, qualitative analyses of audio-taped instructional talk in relation to degrees of scripted instructions during the RCT intervention were examined. The audio-recordings were made in the second and seventh weeks of the program. The recordings were then transcribed and analysed to identify features of teachers' instructional talk. The findings showed patterns of instructional talk being dominated by labelling talk (aligned with the scripted instructions) in the beginning of the intervention and usage of word definitions and extended discourse toward the end of the intervention (aligned with the less scripted components). Significant changes were found only in teachers' instructional talk for the less scripted talk categories.

In paper 3, an evaluation of the factor structure, reliability, construct validity, and criterion validity in the assessment tool Norwegian as a Language for Learning (NSL) was conducted. The NSL tool is commonly used as a measure to identify children in need of L2 support and served as a screener to identity children eligible for the RCT reported in paper 1. The results revealed that NSL was identified as a four-factor model, and measurement invariance was found for language status and gender, establishing construct validity. In addition, significant positive correlations were found between NSL and other standardized language measures (BPVS-II, expressive vocabulary from CELF-4, the vocabulary subtest from WPPSI-IV/WISC-IV, Trog-2 and Bus Story), reflecting criterion validity. NSL can be used as a language assessment tool for students in grades 1 through 4.

Conducting research within the framework of a researcher–practitioner partnership offered opportunities to integrate perspectives from researchers and practitioners to support language learning. The research study used a quality assessment tool to identify eligible students, and the intervention program was beneficial in enhancing children's oral language skills. Including intervention components that varied in the extent to which language instruction was scripted allowed teachers to adapt to students' language skills within the intervention framework.

Acknowledgements

This thesis could not have been carried out without the invaluable contribution of all the children and their parents, principals, the resource teachers who participated in this research study. I am grateful for your hard work and dedication. The intervention was initiated by the former director of education in the municipality, and I offer a special thanks to the former leader and colleagues in the Educational and Psychological Counselling Service. In addition, several people have supported me along the way, and the following deserve my special thanks:

My supervisors and most important collaborators, Vibeke Grøver and Arne Lervåg, thank you for your wisdom, guidance, patience and support. The TextDIM research group, thank you for offering a safe place to present and discuss my research, always accompanied by supportive and constructive feedback. To my colleagues at the Department of Education, and particularly my fellow PhD's Svitlana, Eva, Mari, Kristin, Natalia, Junyi, German, Rebecca, Philip, and Rachelle, thank you for all the lunches, enjoyable company and productive discussions during this period. To Annette, Ulf, Elin, and Olga in the administration stafffor all your help during my Ph.D. Special thanks to Anita, for all the collaborative work, discussions, support, and most of all, a great friendship. To Solveig-Alma Halaas Lyster for her constructive and valuable feedback during the final stages of this thesis.

To David K. Dickinson and Kathy Ganske, for hosting me and making my scholarly visit at Vanderbilt University valuable and productive. They invited me to join their research projects with opportunities to observe, converse and learn first-hand from the participating children and their teachers within a U.S. context.

To Statped for welcoming me and making the transition period in the final stages of this thesis as smooth as possible.

My superb university trio, Astrid, Stine, and Nicolai, for an invaluable friendship. It has been tremendously appreciated throughout these four years.

Finally, my dear family and friends. I am so grateful for having you as my support team. A special thank you to Ole-Vidar, for your love, support, and genuine interest in my research. And most of all for being there for me throughout all the phases of this journey.

Mia C. Heller,
Oslo, February 2020

Table of Contents

PART 1: Extended Abstract

Table of figures and tables

PART II: The Papers

PAPER 1:   Heller, M.C., Lervåg, A., & Grøver, V. (2019). Oral language intervention in Norwegian schools serving young language-minority learners: A randomized trial. *Reading Research Quarterly*, 54(4), 531-552. doi:10.1002/rrq.248

PAPER 2:   Heller, M.C., & Grøver, V. (2019). Teachers' instructional talk in a partly scripted language intervention targeting young language-minority learners: Developments over time. Manuscript submitted for publication.

PAPER 3:   Heller, M.C., & Lervåg, A. (2020). Kartlegging av språkferdigheter for elever på 1.-4.trinn: En vurdering av kartleggingsverktøyet Norsk Som Læringsspråk (NSL). [Assessing students' oral language skills in grades 1-4: An evaluation of the assessment tool Norwegian as a Language for Learning (NSL)].

PART I

Extended Abstract

1   Introduction

1.1 Background

Language skills build the foundation for students' literacy development and facilitate later academic achievement (Dickinson & Porche, 2011; Ellemann, Lindo, Morphy, & Compton, 2009; Storch & Whitehurst, 2002). Support for language learning is particularly important in today's multicultural society considering the increasing diversity in language proficiency among students entering school. Being exposed to language-rich environments can offer unique opportunities for language learning as children participate in interactions where a variety of words are used with meanings derived from the context. Language-rich contexts are also vital for exposure to the language of instruction, especially for language-minority learners (students with language other than Norwegian and Sami) who have not yet reached proficiency in their second language. In school, thus, it is important to understand how students' language experiences can be supported to optimize their learning.

Although studies on language interventions within educational research have led to insight into features needed to enhance students' language learning, an achievement gap is still present in school. Recently, a shift emerged regarding how to develop interventions through researcher–practitioner partnerships to improve the skills needed to succeed in school, "long-term, mutualistic collaborations between practitioners and researchers that are intentionally organized to investigate problems of practice and solutions for improving district outcomes" (Coburn, Penuel, & Geil, 2013, p.2). Research-practitioner partnerships provide additional scope for educational improvement as they strive to merge two perspectives, the researchers' and the practitioners', through iterative processes. Thus, this approach incorporates instructional strategies found effective in research with the needs of practice. One way of promoting student language development in school may be to prepare teachers to interact with students in ways that enhance growth in combination with organizational structures that support learning (learning environment, educational content, teaching methods, and school management). Additionally, assessment tools can provide critical information about the skills acquired, those that are under development, and those that need to be learned.

This research study aims to contribute to the field of language by examining ways to support second-language (L2) learning in the early elementary years. The topic is addressed by evaluating the impacts of an oral language intervention through a researcher-practitioner partnership lens (see Figure 1 by the author of this thesis), combined with analyses of teacher

instructional talk (e.g., modelling, questions, prompts, and comments) and use of assessments to identify students in need of language support, language-minority learners. Illuminating ways to support language learning is a timely topic given the newly published white paper on early identification and inclusive education for children and youth (Ministry of Education and Research, 2019). Students must be relatively fluent speakers to communicate and acquire knowledge. However, some children have not yet reached these levels of proficiency before school entry. Assessing language skills is one of several urgent benchmarks highlighted in this report, showing the importance of early identification of children in need of language support. These benchmarks further address the necessity to develop quality assessment tools to examine L2 skills from kindergarten through high school.
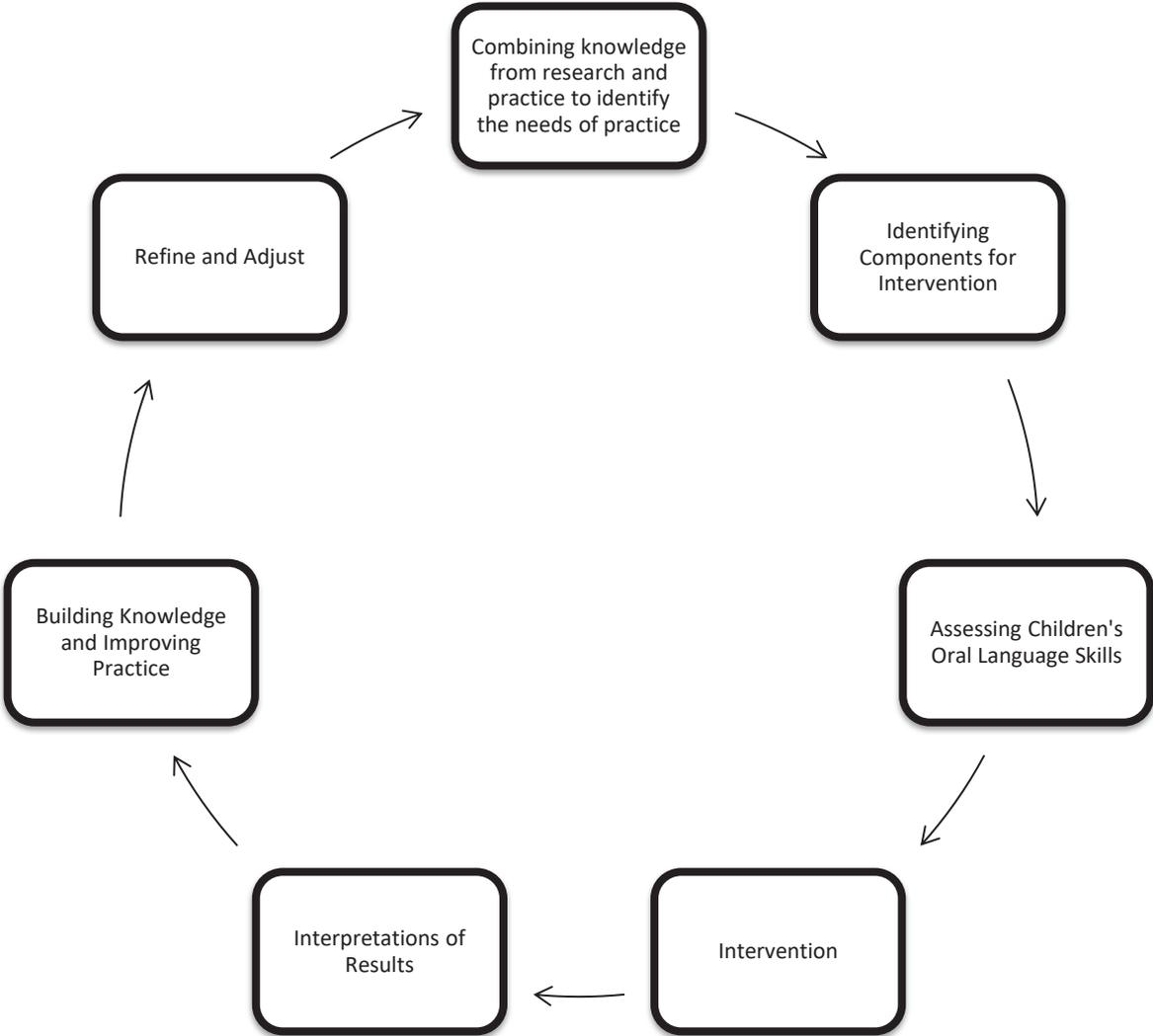
Figure 1. Overview of the thesis' main components through a researcher–practitioner partnership lens.

1.2 Dissertation Outline

This thesis is composed of two parts. The first section is an extended abstract in which the theoretical foundation, methodological considerations, and implications for future research are presented and discussed. The second part constitutes the three papers which is the scientific contribution of this dissertation. Because this thesis builds on a single research study, the first five components of the researcher–practitioner partnership displayed in the model are addressed (Figure 1). The research study begins by combining knowledge from research and practice to identify the needs of practice and ends with interpretations of the results. The first paper lays the ground for the theoretical framing of this thesis, by introducing a social-interactionist perspective to language learning which implies that language develops through participation in interactions with adults inseparable from the social context. The paper also situates the project within researcher–practitioner partnership theory by illuminating ways in which programs can be developed and implemented through collaboration between researchers and practitioners. Paper 2 and paper 3 are situated within these two theoretical perspectives. Whereas paper 2 sheds light on the potential underlying mechanisms in the intervention outlined in paper 1, paper 3 offers insight into assessment as a tool for screening students for eligibility, the quality of such tools, and the ways in which assessment helps to inform interventions and instruction. Taken together, all three papers aim to illuminate the following overarching research question of this dissertation:

How can language-minority learners' oral second-language skills in the early elementary years be supported through a researcher–practitioner partnership?

There are several key concepts in this dissertation, such as oral language, academic language, language-minority learners, language learning, researcher-practitioner partnership, fidelity of implementation, scripting, and assessment. These concepts are presented in the context of their appearance in the text.

2   Instructional Approaches to Language Learning

Oral language is a vital tool for gaining knowledge and fuelling success for students, as it offers opportunities for children to express their knowledge, feelings, and ideas, thus becoming active participants in society. Oral language includes knowledge of words and their underlying concepts, phonology, syntax, and pragmatics. To understand and gain knowledge from the curriculum (both from texts and through classroom talk), some level of academic language skills is needed. Accordingly, academic language is more apparent in students' textbooks compared to the conversations in which they participate. Thus, the written language may be more challenging to comprehend as non-verbal cues (gestures) are unavailable. Academic language has previously been defined as "the specialized language, both oral and written, of academic settings that facilitates communication and thinking about disciplinary content" (Nagy & Townsend, 2012, p. 92). There is an ongoing debate about which skills academic language consists of, in which the emphasis has been on academic vocabulary. To further unpack the underlying mechanisms of academic language, Uccelli, Phillips Galloway, Barr, Meneses, and Dobb (2015) proposed a broader perspective in which the concept is operationalized as "a constellation of the high-utility language skills that corresponds to linguistic features that are prevalent in academic discourse across school content areas and infrequent in colloquial conversations" (p. 338). Academic vocabulary, complex syntax (e.g., embedded clauses) and discourse (e.g., discussions, explanations, arguments, and narratives) are examples of skills required to master content area disciplines, and the amount and quality of input children are exposed to form foundations for acquiring and developing these abstract skills. Additionally, academic language is closely related to students' success in literacy, as these skills aid better understanding of content across subjects (Townsend, Filippini, Collins, & Biancarosa, 2012). Thus, language skills can be a continuum, with conversational talk at one end, and academic talk at the other (Snow & Uccelli, 2009; Uccelli, Demir-Lira, Rowe, Levine, & Goldin-Meadow, 2018).

In the early elementary years, little time is spent on language instruction in classrooms (Biemiller, 2006; Nelson, Dole, Hosp, & Hosp, 2015; Wright & Neuman, 2014), and those who do rarely do so effectively (Carlisle, Kelcey, & Berebitsky, 2013). The amount of time spent on language instruction is particularly disconcerting because language-minority learners

often lack proficiency in the language of instruction (August, Carlo, Dressler, & Snow, 2005). Without such skills children can face obstacles to learning and be unable to access the content (Nagy & Townsend, 2012). Thus, aspects that support language learning and the extent to which children's language skills are easily influenced through interventions should be identified. These aspects are essential for closing the achievement gap and advancing theory and knowledge of language learning in general.

2.1 Theoretical Framing of Language Learning

Multiple theories have illuminated how children acquire language (for an overview, see Hoff, 2014). Throughout this thesis, language learning is situated within the interactionist approach. Various orientations have emerged, and they differ from one another in terms of the emphasis on cognitive processes, interaction patterns, and characteristics of language input (Tomasello, 2003). The social-interactionist-oriented framework guiding this thesis posits that "language develops in the context of achieving pragmatic goals and for the purpose of contact, interaction and knowledge acquisition" (Grøver, Uccelli, Rowe, & Lieven, 2019, p. 2). This perspective builds on Vygotsky's (1986) seminal work, especially how adult interaction assists children's learning and affect its pace. Following this, the overarching principles within the social-interactionist framework include participation in interactions in which the child is given multiple opportunities to use language in exchanges supported by the context in which they take place. The amount and quality of linguistic input accompanied by support from more knowledgeable others are vital for language development. These language experiences also matters for academic learning documented in studies examining how students' knowledge acquisition can be supported through language use, descriptively (Snow, Porche, Harris, & Tabors, 2007; Snow, Tabors, & Dickinson, 2001) and experimentally (Lawrence, Crosson, Paré-Blagoev, & Snow, 2015). Findings revealed close associations between language and literacy skills, and the malleability of language proficiency.

Learning language provides children with a tool kit that promotes knowledge acquisition. As formal schooling begins, students need to expand their oral language skills and acquire school-relevant language skills (advanced vocabulary, grammar, and discourse skills) to learn through language. Cummins (1984, 2001) distinguished between two registers of language proficiency: basic interpersonal communicative skills and cognitive academic-language proficiency. The former register includes everyday language skills used in colloquial conversations (i.e., supported by the context and gestures used by the interlocutor); the latter

register comprises academic language skills that are decontextualized (i.e., separated from the physical context) and thus makes it more challenging to learn. This distinction was based on studies of L2 acquisition that also documented that the academic register needed more years to be acquired. Another influential perspective on language learning was put forward by Halliday (1975), who saw language as a function to create meaning were the child is an active participant in the learning process. According to Halliday, academic language consists of various dimensions that support language-mediated content learning in school. For language-minority learners, this perspective bolsters the challenge of rapidly acquiring L2 skills while simultaneously learning content-knowledge mediated through the same language (Halliday, 1993). Combining Cummins's empirical contribution to the understanding of academic language with Halliday's theoretical-based contribution has previously provided a framework for examining how academic language develops and its underlying components (Snow & Uccelli, 2009). Recent work by Uccelli et al. (2015) replaced this framework with the introduction of a new conceptualization of academic language that was supplemented with the assessment tool Core Academic Language Skills (CALS).

Learning through language builds on children's language competencies. This poses additional challenges for language-minority learners who have not yet mastered the basic, everyday skills of their second language. By incorporating social-interactionist theory as a framework, the research study examines how an oral language intervention potentially promotes students' language learning in elementary school through language exchanges and the contexts in which students participate. The research study builds on language-minority learners' competencies in their second language at school entry and seeks to bridge the gap between basic language skills and academic language skills.

2.2 Features of Linguistic Input Promoting Language Learning

Children's environments and the social interactions in which children participate are fundamental in language learning, and the amount and quality of linguistic input children are exposed to play a pivotal role (Hoff, 2006; Rowe, 2012; Snow, 2014). Researchers have predominantly addressed monolinguals' (L1) language development, with the quantity of talk (i.e., the total number of words spoken) the most influential feature of input in the early stages of language development (Hart & Risley, 1995; Hoff, 2003; Rowe; 2012). Merely exposing children to words, however, is not enough to support language learning over time. As children

accelerate their word learning, the diversity of speech (i.e., variety of words spoken) becomes increasingly important, and this quality of talk including a rich variety of low-frequency words is more influential for language learning than the number of words (Bowers & Vasilyeva, 2011; Dickinson & Porche, 2011; Rowe, 2012; Weizman & Snow, 2001). However, identifying features that support language development goes beyond vocabulary. For instance, exposure to different levels of grammatical complexity support children's comprehension (Huttenlocher, Vasilyeva, Cymerman, & Levine, 2002), which in turn makes it easier to learn new words and gain knowledge. In comparison, decontextualized language use plays a prominent role in preparing children for the discourse used in various settings, especially school. Decontextualized language is characterized as talk that goes beyond the immediate here-and-now, and typically consists of explanations, narratives, and pretend-play over several turns (Snow, 1990; Snow et al., 2001). Extended discourse, "the use of several utterances or turns to build a linguistic structure" (Snow et al., 2001, p.2), is another feature of linguistic input often seen in combination with decontextualized language. In the longitudinal perspective, the amount of decontextualized talk children are exposed to in childhood predicts their language development in school (Dickinson & Porche, 2011; Uccelli et al., 2018). In particular, extended discourse and use of sophisticated words in early childhood have been identified as precursors to academic language (Dickinson & Tabors, 2001; Uccelli et al., 2018).

For L2 learners, research on linguistic input has primarily directed attention to the quantity of talk. However, recent studies expanded this scope by including the quality of the talk, syntactic complexity and extended discourse, often with the teacher the main provider of L2 input to children across age groups (Bowers & Vasilyeva, 2011; Gámez & Lesaux, 2012; Grøver Aukrust, 2007; Rydland, Grøver, & Lawrence, 2014). Bowers and Vasilyeva (2011) examined the effect of teacher talk on pre-schoolers' vocabulary growth over a year. Findings revealed that the amount of teacher talk positively predicted L2 children's vocabulary, while syntax complexity had a negative impact, with the latter being explained by L2 learners' levels of proficiency. Furthermore, word types were not a significant predictor, in contrast to their monolingual peers. In comparison, a study on L2 learners in kindergarten found the amount and diversity of teacher talk are significant predictors of language growth, with the quantity of talk the stronger predictor of the two (Grøver Aukrust, 2007). Moreover, Gámez (2015) revealed that the complexity and diversity of quality input are related to children's language gains in kindergarten, while teacher-dominated talk (the sheer amount of talk) has a

smaller effect on students' language growth. In older students, the impact of linguistic exposure on vocabulary was not related to the amount of talk heard in sixth grade students' vocabulary skills, but the talk quality was (Gámez & Lesaux, 2012). As children enter school, various language skills are needed to acquire school-relevant language. These skills include mastering a diverse vocabulary, comprehending complex sentences, and participating in school-based discourse. Quantity of talk may, therefore, be insufficient to further improve language growth once a solid platform of oral language skills is acquired. These findings can explain why the amount of talk did not predict students' language growth in sixth grade while diversity of talk did compared to younger students in other studies (Bowers & Vasilyeva, 2011; Grøver Aukrust, 2007). Finally, research comparing L2 learners' language trajectories has identified a significant gap in their oral language skills before school entry, and this discrepancy was still present in fifth grade (Mancilla-Martinez & Lesaux, 2011; Rydland et al., 2014).

Taken together, the results indicate that although the number of words, diversity of words and sentence complexity are qualities of exposure that may predict variability in language growth, differential input depending on diverse children's skills, is imperative. More importantly, teachers' instructional talk may differ depending on the instructional settings (Dickinson, Hofer, Barnes, & Grifenhagen, 2014), which can affect the richness of the linguistic input provided to students. Additionally, when the ratio of teacher to child talk is high, children are offered fewer opportunities to respond or engage in conversations (Dickinson & Porche, 2011; Dickinson & Tabors, 2001; Gámez, 2015). These findings have implications when developing interventions to improve students' language learning.

2.3 Intervention Approaches

A key issue is how to leverage instruction time to optimize learning within intervention approaches. A distinction has often been made between implicit and explicit instructional practices, each viewed as beneficial for bolstering language development (Marulis & Neuman, 2010). Implicit instruction is characterized as learning through naturalist exposure, for example, when children are introduced to words, both familiar and unfamiliar, by listening to stories read, and it has often been the main strategy for vocabulary learning in school settings (Neuman, 2011). Although implicit instructions expose children to words, it may be insufficient for students in need of support. A prerequisite for word learning through implicit instruction approaches relies on a solid foundation of word knowledge. Thus, use of explicit

instruction has emerged over the years and become a key ingredient in promoting language learning in interventions within childhood education (Marulis & Neuman, 2010). Explicit instruction is characterized by modeling and scaffolding word meanings and definitions accompanied by materials, and multiple opportunities to practice new words and knowledge with feedback from teachers and peers (Coyne et al., 2019).

Furthermore, instructional approaches are often distinguished as either broad or narrow. Targeting vocabulary can be considered a narrow approach (Marulis & Neuman, 2010); broad approaches are those combining various language skills such as listening skills, vocabulary, sentences, and narratives (Fricke, Bowyer-Crane, Haley, Hulme, & Snowling, 2013; Rogde, Melby-Lervåg, & Lervåg, 2016). There is abundant evidence of the impact of interventions on children's language learning (Coyne et al., 2010; Fricke et al., 2017; Hagen, Melby-Lervåg, & Lervåg, 2017; Nielsen & Friesen, 2012; Silverman, 2007; Silverman et al., 2017; Wasik & Hindman, 2018). However, significant variations in effects are seen with near transfer of skills (custom measures) easier to attain than distal transfer (standardized, global measures; Elleman et al., 2009; Marulis & Neuman, 2010). These differences, in effect, may relate to the intervention design, scope (narrow or broad), dosage (lessons, duration), and group size (classroom or groups). In comparison, in a meta-analysis Rogde, Hagen, Melby-Lervåg, & Lervåg (2019) examined intervention effects on standardized measures of children's oral language skills. Small significant effects were found on measures of vocabulary and language comprehension. Furthermore, differences in effect size were related to the program delivery (high quality implementation demonstrated larger effects) and group size (small groups had larger effects).

2.3.1 Explicit instructional approaches to word knowledge

Leveraging instruction time is vital to optimize language learning. A featured question relates to characteristics of quality instruction. At an overarching level, small-group settings allow teachers to differentiate instruction based on students' needs to a larger degree compared to a classroom setting. Children considered at risk, such as language-minority learners, need more explicit and easily comprehendible instruction than their peers (Marulis & Neuman, 2013). At the component level, building vocabulary has been the most prominent feature of explicit language instruction in early childhood education, followed by components such as active listening, grammar, and narrative skills. Merely targeting many words (vocabulary breadth), however, is not enough to build knowledge (vocabulary depth), as children also need to

process and make connections among words (Nagy & Scott, 2000; Snow, 2017; Stahl & Fairbanks, 1986). Recent vocabulary approaches have distinguished between extended (depth) and embedded (breadth) instructions. Extended instruction incorporates discussions about word meanings. Embedded instruction usually contains brief explanations of target words (August, Artzi, Barr, & Francis, 2018; Coyne, McCoach, Loftus, Zipoli, & Kapp, 2009). Although both approaches have bolstered students' language skills, extended instruction benefits students most (August, Artzi, & Barr, 2016).

Word meaning information and multiple encounters with words

Singular encounters with words are rarely enough to acquire the meaning of a specific word, and the amount of exposure needed may vary according to children's language proficiency. Encountering word meanings through multiple exposures has proven effective to support comprehension (Biemiller & Boote, 2006; Marulis & Neuman, 2013). Semantic information about words supports students understanding of how words are interconnected. Combined with enough dosage, semantic information may help children gain knowledge of the concept behind the words. Activities that support such learning include using pictures, books, multimedia, and discussions that provide the information necessary to build knowledge (Neuman, Newman, & Dwyer, 2011; Silverman & Hines, 2009; Wasik & Hindman, 2011). Using pictures and acting out to exemplify word meanings has been an effective approach for learners with low language proficiency, compared to students with well-developed vocabularies (Silverman & Crandell, 2010). The latter group of learners seems to benefit more from semantic information by exemplifying aspects of word meaning through definitions in combination with antonyms and synonyms. These findings do not imply that learners with less developed vocabularies will not benefit from exposure to definitions, but that definitions must be adapted to learners' proficiency levels for the learners to participate in the process of deriving word meaning.

Relationships among words

Understanding how words are related to one another includes building networks of semantic knowledge to support learning (Hadley, Dickinson, Hirsh-Pasek, & Golinkoff, 2019; Nagy & Scott, 2000; Neuman & Dwyer, 2011). Learning words through relationships such as thematic categories (Pollard-Durodola et al., 2011; Wasik & Hindman, 2011) or taxonomical categories (Newman, & Dwyer, 2011; Neuman, Pinkham, & Kaefer, 2015) is an emerging

approach within educational research. Thematic categories refer to words that are not part of the same type but are typically related through events (i.e., summer/swimsuit). In comparison, taxonomical categories are hierarchical classifications of conceptually related categories (i.e., horses, cats and dogs) based on a shared superordinate (i.e., animals). By introducing words in a categorical manner, students are given opportunities to compare words through structures that create building blocks of knowledge.

Shared book reading

Recent studies have shown support for the use of multiple activities to promote language learning (Wasik & Hindman, 2011). One of the most prominent activities has been shared book reading (Dickinson & Tabors, 2001). This approach exposes children to a variety of words as the story is being read and by incorporating conversations about word meaning, before, during, or after the reading session, enabling students to acquire concepts and knowledge. For instance, using definitions and word relations accompanied by read-aloud has proven to be an impactful practice for teachers to adapt in their classroom (Grøver, Rydland, Gustafsson, & Snow, 2020; Silverman et al., 2014). Subsequent studies expanded on book-reading approaches by differentiating between embedded and extended instruction (Coyne et al., 2009), and book reading in combination with play (Dickinson et al., 2019; Hassinger-Das, et al., 2016).

Modeling language use

Knowledge of how language is used within and across contexts is another essential aspect of word learning which is typically acquired through models of usage (Nagy & Scott, 2000). This type of information provides students with pragmatic knowledge, understanding that words can have different meanings depending on the context in which the words occur. Modeling language use also exposes children to a variety of syntactic features which can promote comprehension. Furthermore, integrating knowledge across contexts has been found to support word learning (McKeown & Beck, 2014).

Content information in teacher–student interactions

Although identifying features that support language learning is fundamental, a better understanding of the nuances of information developed in teacher–students' interactions is needed. This relates to the instructional features of the program components and the

continuum of naturalness occurring within the dialogues. A study by Bowne, Yoshikawa, and Snow (2016) is one of few studies in which the type of information children are exposed to has been examined. Results identified two specific features of information deriving from teachers' instructional talk in kindergarten classrooms: conceptual and declarative information. Conceptual information included concrete, descriptive information about the meaning of words through pictures, and enacting words, providing facts in addition to examples of unrelated information of the targeted word. Declarative information comprised formal definitions and examples of word usage.

In sum, given the wide range of language experiences children show when entering school, leveraging instructional time to optimize learning is vital. Evidence-based components that support language learning are multiple encounters with words across contexts, student friendly definitions, categorization of words, contextual support (i.e., pictures, acting out, modeling usage), and opportunities to interact with words through various activities. Which instruction works best for whom may also depend on the types of information provided in instructions compared to students' language proficiency levels. Thus, tailoring instruction to students' needs through interactions that support comprehension and offer opportunities to generalize knowledge across contexts is essential. These interactions may be especially important to acquire academic language and unlock access to the curriculum.

2.3.2 Explicit instruction in academic language skills

Essential components of input in the early elementary years are rich classroom talk, exposure to academic language through texts, writing experiences, and conversations with multiple turns (Phillips Galloway & Lesaux, 2017). Academic language is the language of instruction in school, and it challenges students' comprehension and use of abstract domains. The level of familiarity with academic language varies among children entering school. Although some children have been introduced to academic language (especially through reading books), at home or in kindergarten, others, in particular L2 learners, encounter such language for the first time in school (Aarts, Demir, & Vallen; 2011; Leseman, Scheele, Mayo, & Messer, 2007). Although young language-minority learners are faced with academic language inputs daily in school, they are not necessarily able to comprehend large amounts of input based on their L2 skill level. Teachers also have been found to simplify their language with L2 learners (Aarts, Demir-Vegter, Kurvers, & Henrichs, 2016). Thus, such simplifications of language may be at the expense of learning. Given the variation in children's access to rich learning

environments before school entry, the importance of explicit instruction in features of academic language becomes even more prominent (Snow, 2014).

Studies on academic language have, to a large extent, targeted academic vocabulary as the dominant feature of emphasis (Carlo et al., 2004). Subsequent studies that incorporated skills beyond academic vocabulary have proven successful (Proctor, Silverman, Harring, Jones, & Hartranft, 2019; Snow, Lawrence, & White, 2009). For instance, a recent synthesis on academic language identified prominent instructional practices in which the outcomes for students' academic achievement across subjects in grades four through eight were improved by using explicit instructions in vocabulary and morpheme, multimodal reading instruction, and strategy instruction for writing (Truckenmiller, Park, Dabo, & Newton, 2019). These results lend further support to academic language malleability, which, in turn, support the importance of explicit instruction.

In sum, engaging in language-rich environments in school is pivotal for developing academic language skills, as it provides opportunities to encounter words that support meaning crucial for developing knowledge (Carlo et al., 2004; Gámez, 2015). Thus, identifying key aspects of teachers' instructional practice is important to bolster language learning across school settings.

2.4 Implementation of Interventions

Fidelity is commonly referred to as the degree to which an intervention is carried out as intended (Dane & Schneider, 1998; Moncher & Prinz, 1991). Fidelity is regarded as a central feature for implementation, as it contributes information for determining the success or failure of interventions (Dane & Schneider, 1998; Durlak, 2010; Mowbray, Holter, Teague, & Bybee, 2003). The main purpose of intervention studies is to examine whether the program will lead to enhancement of the targeted skills. Within early childhood education, numerous programs have revealed positive impacts on increasing support for learning. However, the extent to which fidelity in intervention studies is reported varies, with Swanson, Wanzek, Haring, Ciullo, and McCulley (2011) finding less than half of the studies examined documented fidelity to some degree. Conversly, Hill and Erickson (2019) found 80 % of the studies in their review reported some aspects of fidelity. The difference found may reflect the design inclusion criteria; Swanson et al. (2011) included a variety of intervention designs, but Hill and Erickson (2019) examined only randomized controlled trials. By examining implementation fidelity, underlying mechanisms of change can be identified that inform

teachers' uptake of the targeted components, and to what extent these active ingredients of the program result in the development of students' skills. In comparison, if implementation of a program results in unsuccessful outcomes, fidelity can help identify whether this was due to features related to the design of the program itself, or whether the program was not implemented as intended (Durlak & DuPre, 2008). The former requires a redesign of the program, but the latter may involve measures taken to strengthen the training of implementers in combination with adjustments of program materials. Fidelity levels may, in turn, have consequences for scaling up interventions to larger school settings, as positive outcomes indicate programs' suitability across contexts. Thus, a program's viability in educational settings is largely dependent on implementation fidelity.

2.4.1 Dimensions of fidelity

The increased attention to the quality of support provided in interventions has led to several reviews illuminating how the concept fidelity is defined, conceptualized, and measured (Dane & Schneider, 1998; O'Donnell, 2008). Various operationalizations of the construct fidelity have been proposed over the years, and consensus on a unified definition of the term is still missing. One influential approach to conceptualizing fidelity is the five-dimensional model launched by Dane and Schneider (1998) which consists of adherence, exposure, participant responsiveness, program differentiation, and quality. Adherence concerns the extent to which core components of the program were delivered as intended. Exposure addresses dosage, how many lessons and the amount of time spent on each lesson during the intervention. Participant responsiveness reflects the level of attentiveness or activity seen in the learners. Program differentiation concerns the extent to which core components can be distinguished from the control condition. Quality of delivery refers to how well the teacher uses the strategies, materials, and methods prescribed.

Within educational research, a common approach to fidelity has been to split Dane and Schneider's (1998) five-dimensional model in two, comprising structural fidelity and process fidelity (Mowbray et al., 2003). Whereas structural fidelity applies to the implementation of core components (adherence and dosage), process fidelity illuminates how well the intervention was delivered (i.e., examining the interaction patterns between a teacher and students during a lesson). Within this approach, participant responsiveness is an integral part in both categories. Although separately categorized, these dimensions of fidelity are not dichotomous. Furthermore, the association among aspects of implementation fidelity varies.

14

Odom and colleagues (2010) found high correlations between dosage and quality, whereas Hamre et al. (2010) did not obtain significant correlations between quality and dosage in their study. Inconsistent findings may reflect differences in terms of how each feature was operationalized and measured, in addition to the complexity of the components within each program and the degree of implementation achieved (Durlak, 2010). However, utilization of structural and process features of implementation fidelity is increasing (Mowbray et al., 2003; Odom et al., 2010), and following this approach may provide a nuanced assessment of fidelity, and how it affects outcomes. For instance, Neugebauer, Coyne, McCoach, and Ware (2017) used this multidimensional approach to examine fidelity of implementation and its relation to students' outcomes. By incorporating the quality dimension of fidelity, the authors were able to identify adherence to the program's components in addition to how well it was conducted which, in turn, impacted outcomes in different ways.

2.4.2 Measuring fidelity

The question of acceptable levels of fidelity has brought about debate over the features suitable as quality measures of implementation. It has been assumed that high levels of fidelity are equivalent to improved outcomes, while lower levels of fidelity result in poorer outcomes (Durlak & DuPre, 2008). Although high degrees of structural fidelity indicate interventions are implemented as intended, quality in this sense does not reflect how well the program was executed. For instance, a study revealed that although teachers adhered to a program's components, the quality of instruction (process fidelity) was low (Justice, Mashburn, Hamre, & Pianta, 2008). Conversely, a recent literature review of fidelity in randomized controlled trials indicated moderate and strong levels of fidelity were equally beneficial for positive program outcomes (Hill & Erickson, 2019). To further unpack the differential findings, program complexity and the number of targeted components to be carried out should be addressed. Moreover, the impacts of structural fidelity on students' outcomes have revealed mixed findings (Hamre et al., 2010). These inconsistencies may reflect a mismatch between the program and the context in which the program is implemented, but they may also imply that structural fidelity is less suitable for predicting the quality of an implementation (Piasta, Justice, McGinty, Mashburn, & Slocum, 2015). For instance, Mendive, Weiland, Yoshikawa, and Snow (2016) found significant impacts of an intervention program on adherence and dosage, but only small effects on students' outcomes. The preference for structural measures over process measures of fidelity in educational

research may relate to cost-effectiveness (Harn, Parisi, & Stoolmiller, 2013). Structural fidelity is easy to conduct with, for example, checklists, but process fidelity requires in-depth analysis of what is going on, and thus, is more time-consuming and difficult to measure. In addition, differential associations have been found when structural fidelity and process fidelity are compared with students' outcomes, with literacy best predicted by process fidelity (Odom et al., 2010). Considering the dynamic features of language, process fidelity can provide insight into features of linguistic input in teacher-student interactions. In a study by Silverman and colleagues (2014), the relationship between teacher instructional talk and student outcomes in upper elementary school (third to fifth grade) was investigated. The categories of teachers' vocabulary instructions were identified as: definitions, word relations, application across contexts, morphosyntax and context clues. Comprehension instruction comprised the following categories: literal comprehension, inferential comprehension, comprehension strategies, features of text, decoding and fluency. Results revealed the positive impact of vocabulary instructions on students' outcomes, and inferential instruction on students' comprehension.

Considering the ongoing fidelity debate, several guidelines for reporting fidelity have been posited (Gersten et al., 2005; O'Donnell, 2008). For instance, Gersten and colleagues (2005) put forward a set of indicators in which procedural aspects of fidelity implementation were highlighted with an emphasis on multiple observations of the intervention, using protocols when reporting the number of lessons that have been carried out, and documenting interrater reliability among observers. Quality of implementation was also included by suggesting examination of the varying degrees of integrity. These guidelines are still valuable, as reflected in a current synthesis of implementation (Capin, Walker, Vaughn, & Wanzek, 2018), and usable when examining the extent to which an intervention is used as intended in practice. In practice, the main considerations are usually observational features and clarity regarding the intervention components. Other considerations include differentiation based on teachers' experiences during professional development, coaching support before and during implementation, and to what extent the intervention reflects the context in which the intervention is implemented (Harn et al., 2013).

First, time points chosen for collecting data, such as early or late assessment or a combination of the two, are essential. By conducting early observations, low initial implementation may occur as teachers are in the process of getting to know the program, and sometimes, the students being taught. However, the advantage of an early observation of

implementation fidelity is the opportunity to provide feedback to teachers, while gaining knowledge of the implementation process based on the implementers' responses. If implemented well, this type of feedback ensures that teachers are on the right path, while poor execution can offer teachers much needed feedback on how to go about different components to achieve the program's aim. This knowledge may lead to higher implementation rates by the end of the intervention. Response from teachers can also provide researchers with valuable insight for future professional development and modifications of the program.

Second, the overall quality of implementation fidelity is largely dependent on implementers' understanding of how the program works. The extent to which core components have been explicitly identified can affect implementation quality (Durlak & DuPre, 2008). Teachers' experience and knowledge of the targeted skills can affect implementation: Less experienced teachers are usually more dependent on step-by-step instruction (i.e., specified teacher wording to implement the practices), compared to more experienced teachers (i.e., might need to understand the principles). Furthermore, intervention programs vary in relation to levels of scripting. Scripting refers to the degree to which teachers are required to follow a prescribed manual. It ranges from highly scripted (i.e., reading out instruction word by word) to minimally scripted (i.e., general guidelines to follow but no specified teacher wording) (see paper 2 for an elaborated description). Offering teachers intervention instruction with components with varying scripting may allow teachers to differentiate based on students' needs. The differentiation seen in teacher uptake may also be related to educational cultures across countries. For example, variations in educational quality may occur within and across countries (Bleses et al., 2018; Grøver et al., 2020). Scandinavian countries are characterized by a high-quality universal educational system that values child-oriented learning activities, whereas educational systems in, for example, the United States have shown greater variation and place more emphasis on teacher-oriented learning activities. The degree of implementation fidelity may also reflect how well the intervention program resonates with teachers' needs (i.e., similarity with learning activities found useful in practice). The latter may predict adherence to delivering the intervention as intended, and how well it reflects or stands in contrast to teachers' opinions. Studies have reported that teachers' practices moderate fidelity levels (Gersten et al., 2005), and the extent to which teacher characteristics (i.e., education and teaching experience) predict the fidelity of the implementation was examined in a study by Phillips, Ingrole, Burris, and Tabulda (2017). Findings revealed a significant impact of teachers' preparedness and classroom management

on the quality of the implementation, whereas teacher education was not a significant predictor.

Finally, measuring fidelity also involves assessing the intervention's match to the context in which the intervention is implemented. For instance, it has been underscored that flexibility, such as adaptations fitting a specific context, should be addressed when developing programs (Harn et al., 2013). Flexibility concerns the extent to which teachers are provided opportunities to adapt instructions and possible implications resulting from these adaptations. In a study by Troyer (2019), the degrees and types of teachers' adaptations were examined. Results revealed that adaptation was the most prominent feature of teacher instruction during the implementation, although high variation occurred among the teachers. In this study, only 14 % of the instruction time was not adapted to some extent. Additionally, other studies have found positive effects on students' outcomes when teachers could make adaptations supported by guidance within specific parts of the programs (Kim et al., 2017; Neugebauer et al., 2017). A study by Kim et al. (2017) is one example of how such adaptations can be implemented successfully by teachers with guidance and support from the research team.

Given that the overarching goal of implementing an intervention is to improve students' outcomes, addressing which features of implementation fidelity best predict students' outcomes is important. Following this, intervention studies should consider ways in which teachers can focus on students' needs without deviating from the intervention framework. Examining teachers' practices can also provide much needed insight into how to sustain fidelity. Thus, the ways in which teachers are trained before implementation become an emerging issue.

2.4.3   Professional development to support implementation and outcomes

Teachers' knowledge is closely tied to instruction and child outcomes. Recent work by Schachter, Spear, Piasta, Justice, and Logan (2016) illuminated the relationship between teachers' knowledge and instruction by examining two types of knowledge: disciplinary-related content knowledge and general knowledge of language and literacy development for use in teaching and associations with instruction. Their findings showed that teachers' content knowledge predicted the amount of time spent on language and literacy instruction. In comparison teachers' knowledge of literacy and language predicted students' gains in print knowledge and expressive vocabulary (Cash, Cabell, Hamre, Decoster, & Pianta (2015). Similarly, Wasik and Hindman (2011) found improvements in teachers' quality of

instructional talk (e.g., modeling language and providing children opportunities to use language) were associated with children's gains. In addition, Piasta, Connor, Fishman, and Morrison (2009) showed that teachers' specialized knowledge of language and early literacy in combination with the amount of explicit reading instruction predicted students' gains. These results converge with teachers reporting how knowledge influences their practice. Components such as how children learn combined with information about students' background, their learning goals and school context underlie teachers' pedagogical reasoning (Schachter, 2017).

Studies have addressed how to effectively promote teachers' instructional talk by offering professional development to improve instructions. These studies built on the idea that improving teachers' knowledge and practices will, in turn, lead to quality in instruction and improvements in children's skills (Darling-Hammond et al., 2005). This support is often provided as coaching, workshops, or a combination of the two. Contradictory results were found in recent reviews: Zaslow et al. (2010) reported positive impacts of professional development on teacher knowledge, but Markussen-Brown et al. (2017) did not find professional development affect teacher knowledge. These inconsistencies may be related to differential approaches to professional development (i.e., amount of coaching provided over a period), content, the sample size, and the variety of studies used in the reviews. Additionally, some argue that workshops alone are not an efficient approach for change to occur in teachers' knowledge and practices, as change takes time, and follow-up is needed to support the change (Zaslow et al., 2010). In comparison, Justice et al. (2008) revealed high adherence to intervention components after two days of professional development training in a program targeting language and literacy instruction. However, the amount of professional development needed to implement an intervention depends on the complexity and number of components in the program. This finding is congruent with Markussen-Brown et al.'s (2017) finding that the impacts of professional development on adherence to implementation and medium impacts for quality of instruction. A high degree of structural fidelity was also found in the intervention reported on in paper 1 (adherence and exposure to intervention components), whereas paper 2 reported high levels of process fidelity of implementation (measured as the quality of teacher–student interactions).

Coaching approaches in interventions vary in the extent to which professional development is used to assist teachers with implementation of a supplemental, scripted curriculum (Bierman et al., 2008; Lonigan, Farver, Phillips, & Clandy-Menchetti, 2011), or

helping teachers translate research-based knowledge and related strategies for use in their classroom through interactive talk (Gersten, Dimino, Jayanthi, Kim, & Santoro, 2010; Jackson et al., 2007; Milburn, Girolametto, Weitzman, & Greenberg, 2014). For example, the Literacy Environment Enrichment Program (LEEP: Agder, Hoyle, & Dickinson, 2004; Dickinson & Caswell, 2007) in which the aim was to promote teachers' knowledge and how it relates to improvements in practices by delivering 45 hours of training during a school year is one example of the latter approach. The program had positive effects as the participating teachers improved the quantity and quality of their teaching practices such as organization, use of materials and instructional talk viewed as beneficial for students' learning (Dickinson & Caswell, 2007). Similar findings were found in studies by Neuman and Cunningham (2009), Landry, Anthony, Swank, and Monique-Bailey (2009), and Jayanthi et al. (2018). In contrast, Wasik, Bond, and Hindman (2006) expanded the ingredients previously used in professional development interventions by conducting one-on-one coaching in combination with group workshops. This support included conceptual and practical knowledge. Teachers were also trained in book reading and oral language strategies with scripted lesson plans for implementation of a supplemental curriculum. Findings revealed positive effects on teachers' use of language-promoting strategies.

Exploring the mechanisms underlying change in teacher practices is important for educational improvement. Although the amount, intensity and type of support offered varied, the studies reviewed showed the importance of supporting teachers in how they can adapt instructional talk based on students' needs and classroom contexts. These aspects affect how fidelity is measured and the assessments used in future studies.

2.5 Assessment

2.5.1 The purpose of assessment

Identifying and monitoring students' development in educational research and practice relies on assessment. Assessment offers an avenue for understanding children's competencies and resources, which within the educational context mainly concerns examining students' skills in relation to implications for academic achievement. A topic within educational research targeting pull-out interventions relates to identification of children in need of intervention, utilizing results to improve instruction and prevent difficulties in students' development.

The purpose of assessment can be split into four sections: screening for identification, establishing baseline skills, identifying components for use in interventions, and examining effects of programs (Sattler, 2018). Screening is typically used for early identification as it taps into a set of skills representatives of the targeted domain of interest. Screening is commonly used to identify students eligible for intervention. Assessing baseline skills is considered a broader examination than screening, because it seeks to establish an overview of a child's general performance across a broad set of skills within the targeted domain. To identify components and related procedures when developing interventions, assessment is used to distinguish between skills mastered and those in need of intervention. Examining effects utilizes assessment to compare students' results at specific time points (see paper 1).

Based on the specific purpose, the assessment may have different functions depending on whether it is used to examine individual performance, class or school performance, or a country's performance or is used as a political instrument to guide decisions concerning the educational platform put forward by policy makers. For instance, schools rely on assessment to inform students' progress and improve instruction. Assessment also plays a pivotal role when selecting students for participation in interventions and examining potential effects of programs in educational research. In comparison, assessment has been utilized to inform decision making, such as educational goals and reforms by policy makers. Although differences can be seen across educational research, practice, and politics, the intent, responsibility and usage of assessment still build on the same foundation.

2.5.2 Theoretical framework for language assessment

Language is a complex system composed of receptive (listening) and expressive (speaking) skills and semantic, phonological, pragmatic and grammatical structures. Accordingly, to assess the degree to which students have acquired a specific set of language skills, characteristics of typical and atypical development trajectories are needed. Knowledge of when specific language skills develop and to what degree and age mastery should be expected is vital. This relates to the biological, cognitive, and behavioral aspects of children's development in combination with environmental influences (Hulme & Snowling, 2009). Additionally, longitudinal studies and intervention studies in early childhood education have contributed to knowledge of language development and learning. These findings continuously underscore the need for early assessment to identify children in need of language support. For instance, competencies in listening and speaking skills may differ at various time points and is

highly influenced by the language environments in which children participate. A common practice to account for variations in language skills within and across age groups has been to develop a set of norms as a reference for interpreting results for individuals and groups of students.

A current issue within educational assessment is that some tests do not meet acceptable standards (i.e., being reliable and valid), yet are used frequently to assess students (Arnesen, Braeken, Ogden, & Melby-Lervåg, 2019). Examining students' language skills is largely dependent on the quality of the tool and the assessment competence of those who conduct and interpret the results. Information derived from assessments can inform intervention in several ways when interpreted correctly. First, assessment identify which students need additional support. Second, skills that are not mastered are revealed through assessment, which makes it easier to decide on components that should be targeted during intervention. Conversely, inappropriate interpretations of results may prevent children from receiving language support. Furthermore, the extent to which assessment results are utilized to inform instruction varies, which is reflected in assessment practices found within and across school settings (Arnesen et al., 2019). Teachers are mainly responsible for conducting these assessments, and preference for assessment tools may largely depend on time effectiveness and experience.

2.6 Summary

Situating this research study within a social-interactionist approach demonstrate the vital role of interaction and the context in which it takes place to facilitate language learning. This involves exposing language-minority learners to language-rich environments where multiple encounters with oral language skills are embedded through interactions with peers and adults in various contexts. Given the lack of explicit language instruction in school, this lens may help inform ways to support language learning more efficiently in the early elementary school years. Furthermore, identifying students in need of language support, such as language-minority learners, depends largely on high-quality assessment. Assessment can provide information about students in need of language support and the type of skills in need of enhancement.

3 Methodological Perspectives and Considerations

3.1 A Researcher–Practitioner Partnership Framework

Intervention research within childhood education has commonly been dominated by a translation-to-practice approach, a power line from researchers to practitioners, yielding insights through a top-down lens. This model undermines the knowledge and expertise in the field of practice, by seeing educators as consumers of knowledge (Gutiérrez & Penuel, 2014). The emergence of researcher–practitioner partnership interventions reflects an important shift within educational research. It has moved away from the unidirectional approach to an approach characterized by colliding perspectives. Researcher–practitioner partnerships converge the perspectives of researchers and practitioners, acknowledging the competence and knowledge both parties bring to the partnership (Coburn et al, 2013; Snow, 2015). Thus, the overarching principle is the partnership, with features of structure, support, and sustained collaboration guiding the process. Questions resulting from researcher–practitioner partnership approaches expand from what works to include how practice and research can inform one another. Thus, the focus is not just on solving problems but also on building knowledge through collaboration which can impact practice.

3.1.1 Researcher–practitioner partnership approaches

There are three types of researcher–practitioner partnership approaches: research alliances, design research, and networked improvement communities (Coburn et al., 2013). Research alliances are collaborations between organizations and districts that focus on solving local issues within or across sectors to inform policy and practice. The participants' roles are distinct throughout the partnership, and collaboration is restricted to specific periods of the research process. Design research attempts to enhance instruction in education while simultaneously informing theory by constructing and examining curricular activities and materials. In contrast to research alliances, design research values collaboration among participants during every phase of the project. Networked improved communities concentrate on solving problems that are common for multiple communities. Solving problems involve drawing on shared experience from various settings to illuminate what works where, when and under what conditions, in particular, reversing the roles of practitioners and researchers

within the partnership. Practitioners perform the data collection and analysis, whereas researchers mainly facilitate the ongoing process (Coburn et al., 2013). The present study is situated within a design research partnership, seeks to improve instructional practices for L2 learning and inform theory by developing and studying curricular-related material through iterative processes between researchers and practitioners.

Within researcher–practitioner partnership interventions, the main features are the emphasis on collaborative processes and implementation, targeting of practical problems, and relevance of the local contexts (Gutiérrez & Penuel, 2014). These features stand in contrast to researcher-developed interventions where researchers posit the aim of the study, often solely based on a theoretical point of view. The value of iterative processes is reflected in an influential language intervention emerging from a researcher–practitioner partnership, the Word Generation program (Snow et al., 2009). Researchers and practitioners' collaboration processes resulted in strategies and activities that met the needs of teachers. Word Generation revealed significant impacts on students' language growth, with larger gains for language-minority learners. The following illuminates the iterative processes within the researcher–practitioner partnership in present research study.

Iterative processes: identifying the practice need

The iterative processes in this research study combined knowledge from practitioners (school psychologists, principals, teachers, and the department of education) and researchers which formed the foundation of the partnership. The partnership identified the practice need as finding ways to solve current challenges with L2 instruction. The main issues concerned the instruction content, instructional practices and organization. For instance, the organizational framework of L2 support in Norwegian elementary schools is characterized by flexibility, such as small groups or an assistant in the classroom. Additionally, the amount of instruction and the use of materials vary considerably. Acquiring proficiency in L2 skills takes time. Features like the amount and quality of linguistic input combined with opportunities for students to use their language play a vital role in students' language learning (Gámez, 2015; Grøver Aukrust, 2007). Considering this, the partnership decided on a framework comprising eight 30-minute lessons per week over eight consecutive weeks to bolster students' L2 learning. Five of these lessons were morning lessons, whereas the remaining three were afternoon lessons. Given the importance of learning language through interaction with more knowledgeable others (Grøver et al., 2019), the groups consisted of four to six students and

one teacher. Thus, the dosage and the group size provided the initial framework for constructing the intervention program.

Iterative processes: the relevance of the local context

Deciding on core language components and the content in each lesson addressed the relevance of the local context. These processes included a field trip to schools in suburban districts of London, observations across first grade classrooms, conversations with first grade teachers, multiple meetings between researchers and practitioners, and pilot and revised processes of the intervention program. Combining knowledge from the field trip to schools in London with information from meetings within the partnership illuminated issues related to instruction and material that supports L2 learning in the early elementary years. For instance, what works in one educational context may not be efficient across educational contexts. Variations across contexts reflect another key feature of the researcher–practitioner partnership, the relevance of the local context when constructing and implementing interventions (Coburn et al., 2013). The partnership then conducted several observations across first-grade classrooms in various schools. During the observations, the targets were the words, sentence structures, and instruction practices used by teachers, interaction patterns between the teacher and students, engagement among students, and the materials used. Conducting these observations also resulted in conversations with teachers about the instruction content and daily challenges when planning instruction that support L2 learning. It became clear that time- and cost-effective material was a necessity in the intervention program given the minimal preparation time available for teachers to make instruction material themselves. This information helped when discussing and deciding on core components, combined with findings from evidence-based research. Teaching vocabulary and basic sentences accompanied by invitations to participate in extended discourse became the core components in the intervention program, all supported by visual material.

Numerous studies have reported on effective strategies for word selection and instruction to improve children's language skills (Beck & McKeown, 2007; Biemiller, 2005). The partnership used this knowledge to form the basis for word selection. Additionally, iterative processes revealed that basic content words (i.e., nouns, verbs, and adjectives used in everyday conversations) are not explicitly taught in Norwegian first grade classrooms. This relates to the assumption that students acquire these words before school entry. However, these are words language-minority learners need to build their foundation of L2 skills. Thus,

the partnership decided to target basic content words needed to get access to the curriculum. The program also used a four-level structure (each with a two-week duration) to gradually increase the level of difficulty based on the initial L2 competencies of the language-minority learners. Together, these four levels comprised 19 superordinate categories and 300 words (see paper 1 for details about word selection and categories).

Iterative processes: implementation

Implementation reflects how well an intervention is carried out. Implementation is a critical feature when developing and piloting intervention programs, in researcher–practitioner partnership interventions and in researcher-developed interventions. Researcher–practitioner partnerships highlight implementation as a core feature and offer ways to shed light on what is going on in the program through direct observation and feedback from implementers. The latter implies "knowing what aspects of a new program or practice are easy or hard to implement, which ones are adopted after minimal versus only after intensive professional development, which are embraced by teachers, and which rejected is crucial to designing new innovations that are likely to take" (Snow, 2015, p. 461). The intervention program was piloted two times. In both pilots, teachers received professional development as a one-day workshop followed by two weeks of preparation time. This workshop revealed large variations in teachers' knowledge of language development and strategies for supporting language learning. Conducting multiple observations during the implementation of both pilots offered an avenue for evaluating the teachers' uptake, while providing and receiving feedback. The observations also enlightened about how students responded to instruction and the various components chosen to support their language learning. Evaluation meetings with teachers and their leaders, halfway through and after the implementation period, provided information about the extent to which the program met the teachers' needs. For instance, teachers reported awareness and value of teaching word meanings through relationships and multiple encounters with words across activities. In addition, lesson structure and the number of activities per lesson were modified in accordance with teachers' feedback. Knowledge gained from these iterative processes informed the final revisions made to the intervention program and is something researcher-developed interventions miss out on (Donovan & Snow, 2018). These qualitative aspects of implementation are needed to elucidate the relevance and utility of research to increase knowledge and understanding for policy and educational improvement (Snow, 2015). Thus, whereas previous evaluations of implementation have been

26

regarded as hierarchical and summative, researcher–practitioner partnerships focus on formative and collaborative monitoring (Power et al., 2005).

In sum, the researcher-practitioner partnership intervention in this thesis sought to find ways to bridge the knowledge gap for language-minority learners. Iterative processes resulted in an overarching framework comprising 64 30-minute lessons across eight consecutive weeks and groups of four to six students. The core components were vocabulary, basic syntax, and extended discourse, all supported by visual material. These components appeared across multiple activities (labeling pictures, concept maps, drawing, games, songs, and exposure to language outside the classroom), all proven effective in supporting language learning (Neuman et al., 2011; Silverman & Hines, 2009; Wasik & Hindman, 2011). Furthermore, participating in interactions with more knowledgeable others offered multiple opportunities for the novice learners to understand the meaning of words, syntax, and discourse and practice these skills on an everyday basis. Additionally, how well an intervention captures the outspoken need of practice determines the intervention's viability. The advantages of using a researcher–practitioner partnership framework compared to researcher-developed framework is the emphasis on tackling the problems posited by practitioners, the iterative collaboration processes between researchers and practitioners together with a broad focus on implementation processes. These features may provide a better fit to the context in which the intervention program is implemented.

3.1.2 Design

This research study used an RCT, where participants are randomly allocated to one or two groups to examine the effect of a specific treatment condition. RCT designs are considered the gold standard of research where the effects of an intervention on a specific set of skills are the main interest (Shadish, Cook, & Campbell, 2002). For instance, using a pre-test post-test design with random allocation (paper 1) provides better control of possible differences between groups at baseline. Pre-test scores can be used as a control when analysing intervention effects. These scores often explain much of the variation seen in post-test scores and increase the power to detect differences among the intervention group and the control group. The design chosen in this research study was particularly based on ways to reduce threats to internal validity (i.e., drawing inferences from the obtained effect). For instance, the RCT design randomly allocates participants into an intervention group and a control group which reduces selection bias, as differences in characteristics can be assumed to be randomly similar in each group. The randomization then reduces the risk of overestimating or

underestimating the intervention effect. Conducting research with children as participants who display a rapid development trajectory of school relevant skills is challenging. By randomly allocating children into groups, the RCT design controls for changes related to maturing, regressing, or history during the intervention period. However, threats such as attrition (losing participants due to relocation) and the contamination effect (i.e., treatment diffusion) cannot be controlled for by the RCT design. In the present study, only eight children were lost between the post-test and four-month follow-up post-test. Attempts were made to locate students who relocated to neighbouring schools within the municipality. In addition, resource teachers implemented the intervention. Possible threats to treatment diffusion may be weakened by this decision, as none of the participating resource teachers were the students' classroom teachers.

Given that the effects of an intervention are the main interest when RCT designs are used, the extent to which mechanisms supporting this effect are explored varies. This is why many intervention studies are referred to as "the black box", implying that the underlying mechanisms of the observed effect are unknown. The advantages of supplementing an RCT design with a qualitative feature relates to the possibility of addressing what is going on during the intervention lessons (process fidelity). Such analysis can offer knowledge of teacher uptake and variations among teachers, and generate hypotheses that can be tested in future RCTs. Thus, conducting RCTs within a researcher–practitioner partnership can provide a different pathway to how interventions are perceived in practice as the partnership moves away from a strict top-down process. The collaboration between the two fields results in a focus on the main concerns of practice when developing a program for implementation in education. The researcher–practitioner partnership also differs in relation to how implementation is monitored and interpreted. Furthermore, complementing an RCT design with materials suitable for qualitative analysis (i.e., transcripts of audiotaped recordings) afford examinations of the interactions taking place between the teacher and the students during the intervention lessons. Information derived from these processes is needed to shed light on ways to improve instructions that supports language learning in school.

3.2 Validity in Language Assessments

Language assessment takes its form through theory and operationalization of the construct. This lays the foundation for constructing assessment tools and interpreting results considering the underlying theory (Thorndike & Thorndike-Christ, 2014). Children's language

development is a cumulative process, and skills take different forms across ages. Variations in language skills and developmental trajectories (i.e., when a specific skill is acquired, and the level of proficiency expected within age groups) pose challenges when choosing measures for assessing students' language skills. In addition, by turning a language construct into an observable unit (i.e., an assessment tool), constraints that threaten construct validity may occur. Construct validity relates to the congruence between the theoretical term and the operationalized term. Constraints refer to "noise" in the forms of systematic error (the degree to which a construct is underrepresented) and random error (construct irrelevant variance), which can lead to biased results (Messick, 1995). The latter concerns reliability, the ability to produce consistent results within similar conditions (Shadish et al., 2002). A precondition for validity is the reliability of an assessment scores. In the present study, language assessments commonly used in language intervention studies (e.g., Grøver et al., 2020; Hagen et al., 2017; Rogde et al., 2016), served as indicators to demonstrate consistent measurement and interpretation of the language construct. Cronbach's alpha revealed high internal consistency in each of these language assessments in the present research study (see paper 1 and paper 3).

A distinction between observed variables and latent variables is made when assessing student's language skills. Observed variables concern the test scores (i.e., language scores), which are comprised of measurement errors and true variance. Latent variables refer to hypothetical constructs (factors), a continuum of a phenomenon that is not directly observable (Kline, 2016). Structural equation modeling (SEM) offers a way to explore theoretical hypotheses, in this case, relationships among language factors. The shared variance of the observed variables that go together in an overarching concept (i.e., language) is interpreted as the true score variance (Brown, 2015). Additionally, SEM examines whether measures reflect the same construct across groups and time through analysis of measurement invariance (Putnik & Bornstein, 2016). This is a precondition for comparing group means, a crucial aspect when assessing language skills across time points and groups in an intervention or examining aspects of quality in language assessment tools. For instance, scalar invariance between pre-test and immediate post-test scores was found in the intervention study (paper 1). In comparison, findings of scalar invariance for language status and gender in the assessment tool reported on in paper 3 made it possible to compare group means. Thus, the main advantages of latent variables compared to observed variables, is the possibility of testing theoretical hypotheses about language factors without measurement errors which results in more valid conclusions.

3.3 Generalization of Research Findings

Situating research within educational contexts is challenging for many reasons. The setting and the processes occurring within schools and school districts are highly contextualized (i.e., persons, resources, and environment), and the ways in which control of variables can be managed is demanding. These factors may affect generalization of research findings, the generalization across persons, treatments, settings and outcomes (Shadish et al., 2002). Figuring out what works, under what condition and for whom reflects possible limitations with generalization of research findings. Thus, utilization of implementation data (structural and process fidelity) is crucial when interpreting outcomes and possible generalization drawn from educational research studies.

Single studies, such as the present intervention study, strive to achieve high external validity through selection procedures by identifying characteristics representative of participants and settings. For instance, the participating schools represented the multicultural population and socioeconomic status of the municipality. Language-minority learners are a heterogeneous group in Norway and represented various languages in this study. Because children's test scores were used as inclusion criteria (1.5 standard deviations below the mean or lower on a language test), the inclusion criteria might affect the generalization of the findings. Therefore, the findings cannot be generalized to language-minority learners scoring above this threshold.

3.4 Ethical Considerations

Ethical principles underpin several norms to consider when conducting research (National Committees for Research Ethics in Norway [NESH], 2016). The research study sought to find instructional contributions to support L2 learning in school and was approved by the Norwegian Centre for Research Data (NSD) (Appendix A). In line with the NSD's requirements, all data were stored in a secured place and participants were anonymized. Consent was obtained from the schools' principals, resource teachers, and children's parents (see appendices B–D).

Within an educational context, multiple perspectives on best practices of instruction exist. An ethical concern in this research study related to the expectation of documenting an effect from policymakers and practice. At the school level, this can be viewed as positive by reflecting a genuine interest in gaining knowledge of the beneficial types of instruction to implement. At the policy level, result may provide insight into how to organize language

instruction in the ongoing debate about educational improvements to support language learning. Thus, the need for evidence-based knowledge provides insight into the ongoing debate of how to support L2 language learning in school.

3.4.1 Children as participants

An overarching principle when using children as participants is to continuously consider what is most beneficial for the child at all stages of the research process. Participating students were a group of language-minority learners in the early elementary years with low proficiency levels in their L2. As these students were between five and seven years of age, additional steps were taken to obtain consent. Children as participants have lower consent ability than adults and are more likely to be obedient to authority figures. To obtain parental consent, translations in the most common minority languages within the municipality were provided (Urdu, Somali, Arabic, Kurdish, Turkish, Tamil, Vietnamese, Polish, English, and Albanian). The consent form contained information about the intervention program, allocation of students to an intervention or a waiting-list control group, types of assessment, and procedures for privacy and data protection (Appendix B). Parents also contributed with information about their child's early childhood care, family language use, and parental education through a questionnaire (Appendix E).

Minority groups are at risk for stigmatizing and being categorized as a vulnerable group according to the NESH (2016) guidelines. The language assessment was used as a tool to identity students in need of L2 support and to assess possible effects of the intervention on children's language skills. Ethical applications of assessment use and practices play a vital role in this process, and care was taken in the research study to protect children from additional stigma and provide equal testing situations for all children. For instance, all assessments were conducted at the participating students' schools by the author of this thesis and specific guidelines for testing were followed. Scoring and interpretations of results were examined based on a standard manual which was supplemented with inter-rater agreements. Given that the NSD categorizes language as an ethnic marker, only the frequency of language types in the sample were reported when results were published in peer-reviewed journals. Findings regarding differences among groups were reported only between the intervention group and the waiting-list control group.

3.4.2 The interplay between methodological and ethical considerations

Another ethical issue when conducting educational research relates to the interplay between methodological and ethical considerations. This dilemma is commonly seen in relation to the choice of design (Tangen, 2015). Taking ethical considerations into account when deciding on the preferred design highlights and provides insight into how to protect informants in the best possible way. These decisions also attend to how the research study applies to the ethical standards throughout the process. Additionally, choice of design can increase ethics by protecting human dignity and privacy (NESH, 2016). A randomized controlled trial is the preferred choice of design when examining effects, as it provides control of variables by incorporating random allocation of participants into groups. To explore the effect of a program, it is critical to be able to compare participating students by using an intervention group and a control group. In this research study, the control group was business as usual (teaching within the classroom). Although the RCT design is considered the gold standard in research, it is still subject to ethical dilemmas. One of these relates to whether the intervention will influence children's languages skills compared to the amount of time spent on a supplemental teaching approach. For instance, the intervention may not have the intended effect in which students following the standard instruction have better progress than those participating in the intervention. Another aspect concerns the use of a control group, and the extent to which it is unethical to use randomized sampling to allocate students into an intervention group and a control group. A waiting-list control group design with three test time points was used in the intervention study. All students in the waiting-list control group received the intervention after the post-test was completed. To compare the results across the two groups, the four-month follow-up post test was conducted after the waiting-list control group had received the intervention. Thus, these students did not lose something by being allocated to the control group in this study.

In sum, the interplay between research ethics and the need for research-based knowledge is balanced through the researcher-practitioner partnership framework in this study. This relates to the inclusion of concerns and experiences practitioners bring to the partnership. Many of the practitioners who participate in researcher–practitioner partnership work daily with the targeted children of the research project. Thus, they are responsible for utilizing and sustaining the findings emerging from research in practice.

4 Presentation of Papers

4.1 Paper 1

Title: Oral language intervention in Norwegian schools serving young language-minority learners: A randomized trial. *Reading Research Quarterly*, *54*(4), 531–552. doi:10.1002/rrq.248
Authors: Heller, M.C., Lervåg, A., & Grøver, V.

Objectives

This study presents an RCT in which the aim was to shed light on how structured and specific language instruction in combination with visual material can enhance language skills for children learning an L2. The following research questions were examined: 1) Can young language-minority learners' L2 skills be improved by participating in an eight-week researcher–practitioner partnership oral language intervention?, 2) To what extent are group differences present after the waiting-list control group received the intervention?

Method

Participants

Language-minority learners from first and second grade across 16 schools within the same municipality were invited to participate in this study ($n = 137$, mean age = 6 years 3.34 months, SD = 6.20 months). Eligible students were those scoring 1.5 standard deviation below the mean on the standardized Norwegian language test, NSL.

A waiting-list control group design was used to measure the effect of the program. The waiting-list control group followed business as usual and received the intervention after the intervention group had completed the program. Business as usual was composed of classroom teaching (one teacher and an assistant responsible for 20–25 students) characterized by flexibility in instructional practices. However, literacy instruction in Norwegian elementary schools is based on the phonetics tradition.

Data collection

Children's language skills were assessed with five language measures on three different occasions: before the intervention, immediately after, and four months later (after the waiting-list control group also had received the intervention). Standardized measures were used to

examine receptive (BPVS-II) and expressive vocabulary (Expressive Naming subscale from CELF-4, Definition subscale from WPPSI-IV/WISC-IV), grammar (Trog-2) and narrative skills (Bus Story). All tests were translated into Norwegian and standardized, except the narrative measure Bus Story, which was researcher-translated.

The intervention program

The intervention program was developed in partnership with the researcher and practitioners with varying expertise in L2 learning. The main components were language-rich interactions in small groups (four to six children), where children learned vocabulary skills through categorization of words, combined with basic levels of syntax and discourse skills. A key feature was to teach language-minority learners words not explicitly taught in the classroom with basic communication skills to access the curriculum.

Analysis

Raw scores were used, and all analyses conducted with M*plus* (Muthén & Muthén, 2017), and SEM. The effects of the intervention on students' language skills were estimated by creating a dummy-coded group variable and calculated from the y-standardized coefficients. These coefficients can be interpreted as equivalent to Cohen's d (Brown, 2015).

Results

The results revealed that students in the intervention group demonstrated significant improvements in various oral language skills compared with the waiting-list control group. Additionally, there were no longer statistically significant differences between the groups after the waiting-list control group had received the intervention.

4.2 Paper 2

Title: Teachers' instructional talk in a partly scripted language intervention targeting young language-minority learners: Developments over time. Manuscript submitted for publication. Authors: Heller, M. C., & Grøver, V.

Objectives

The second study examined modifications in teachers' instructional talk (e.g., modelling, questions, prompts, and comments) in relation to components that varied in scriptedness (the

degree to which teachers are required to follow a prescribed manual). The following research questions were examined: 1) What characterized teachers' instructional talk in the early and late parts of the intervention?, 2) Did teachers modify their instructional talk during the intervention?

Method

Participants

The sample consisted of 15 teachers and 65 students (the intervention group) from the randomized controlled trial reported in paper 1.

Data collection

Audio recordings were used to collect data on teacher–students' interactions, and teachers audiotaped three lessons in the beginning (week 2; t1) and the end (week 7; t2) of the intervention. Talk categories were developed that helped identify how different intervention components appeared in teacher talk more specifically, what was identified as scripted intervention component (picture labelling and repeated exposures of targeted words), more softly scripted components (definitions and relationships among words) and finally, minimally scripted components (i.e., discourse moving beyond the here-and-now). Additionally, measures of tokens and types of talk for teachers and the student groups were included as check points when the scripted and less scripted talk categories were examined. These decisions allowed the teachers' discourse utterances to be comprehended more fully.

The intervention program

The program was the researcher-practitioner partnership oral language intervention described in paper 1.

Analysis

Descriptive analyses examined frequency patterns of talk categories at t1 and t2. It was expected that as teachers got to know their students better, changes in teacher instructional talk in relation to scriptedness may occur. To assess modifications in teacher instructional talk, paired sample *t* tests were used. All measures of talk exposure were calculated per minute as lesson duration varied across groups.

Results

The result revealed teachers used high amounts of labeling talk (simple and elaborated labelling) in the beginning of the intervention. Both labeling categories aligned with the scripted components of the intervention. At t2, although teachers still used simple labelling to a large extent, we also found higher usage of extended discourse among teachers. Furthermore, findings revealed significant increases in the talk categories conceptual definitions, usage-based examples of definitions, and extended discourse (talk beyond the immediate here-and-now). These categories aligned with the less scripted components of the intervention. The scripted components (simple and elaborated labeling) did not change significantly, although elaborated labelling approached a significant decrease.

4.3 Paper 3

Title: Kartlegging av språkferdigheter for elever på 1.-4.trinn: En vurdering av kartleggingsverktøyet Norsk Som Læringsspråk (NSL). [Assessing students' oral language skills in grades 1-4: An evaluation of the assessment tool Norwegian as a Language for Learning (NSL)].
Authors: Heller, M. C., & Lervåg, A.

Objectives

The overall aim of this study was to examine the factor-structure, reliability, construct validity and criterion validity of the language test NSL (Frøyen, Ahmadinia, Heller, & Skjåk, 2011; Frøyen, Ahmadinia, Heller, Skjåk, & Namvar, 2015). The following research questions were addressed: 1) Does the NSL show satisfactory internal consistency (reliability)? 2) Does a one-factor structure or a multifactor structure fit the data best (construct validity)? 3) Does the NSL show measurement invariance across language status, gender and grade?, 4) To what extent is the NSL related to other standardized measures of language?

Method

Participants

This study comprised two samples. The first sample was the NSL norm sample from 2014. Students ($n = 373$) from first to fourth grade across 41 schools participated; 60.1 % of the sample were monolingual students (47.5 % girls, mean age 92.52 months, $SA = 14.00$ months). Sample selection reflects the multicultural population in schools. Students were not

eligible for special needs education, and the minority learners in the sample were not eligible for language support in L2. This sample represented a total of 40 minority languages.

The second sample was the students ($n = 137$) from the intervention study in paper 1. In contrast to the NSL norm sample, the RCT sample comprised L2 learners in need of language support. Sample selection identified students scoring 1.5 standard deviations below the mean on the NSL. This sample represented a total of 31 minority languages.

Measures

The NSL assessment tool was comprised 195 items targeting various language skills, such as categorization of words, vocabulary (receptive and expressive), and grammar. The standardized language measures included language skills such as receptive and expressive vocabulary (BPVS-II, Expressive Naming subscale from CELF-4), definitions (Definition subscale from WPPSI-IV/WISC-IV), grammar (TROG-2), and narrative skills (Bus Story).

Analysis

Raw scores were used, and all analyses performed with M*plus* 8 (Muthén & Muthén, 2017). Confirmatory factor analysis examined the NSL's structure and level of measurement invariance (construct validity). Correlational analysis explored the relationship between the NSL and the five language assessments (BPVS-II, Expressive Naming subscale from CELF-4, Definition subscale from WPPSI-IV/WISC-IV, TROG-2, and Bus Story).

Results

The confirmatory factor analysis identified NSL as a four-factor model, establishing construct validity. Evidence of criterion-related validity was found in correlations between the NSL and the five language tests (BPVS-II, Expressive Naming subscale from CELF-4, Definition subscale from WPPSI-IV/WISC-IV, TROG-2, and Bus Story). Thus, the NSL is seen as suitable for assessing monolingual and language-minority learners' language skills in grades one through four.

4.4 Summary

The research study's three papers all aimed at examining ways in which language learning can be supported for language-minority learners in the elementary years. Although intervention effects, modifications in teacher instructional talk and validation of a language

assessment tool was investigated separately in each paper, the topics overlapped and completed each other. The main features of each paper are summarized in Table 1.

Table 1

Summary of common features and key terms among the three papers

	Paper 1	Paper 2	Paper 3
Theme	Oral Language Skills Minority Learners Intervention	Oral Language Skills Minority Learners Instructional Talk	Oral Language Skills Quality of Assessment
Key terms	Researcher–Practitioner Partnership	Scripted, Softly Scripted, and Minimally Scripted Components	Language Components
Assessing	Assessment of Language Skills	Assessment of Talk Categories	Assessment of Components
Assessing	Effect Fidelity of Implementation	Modification Interrater Agreement	Usability Validation
Analysis	Validity and Reliability	Validity and Reliability	Validity and Reliability
Analysis	Descriptives, Regression, SEM	Descriptives, t-tests	Descriptives, SEM

5 Discussion

5.1 Supporting Language Learning in Young Language-Minority Learners

Decades of research highlight how early a knowledge gap occurs among children. This difference is reflected in their oral language skills before transition to school which illuminates the importance of interventions. Intervention studies on language learning inform our understanding of strategies that can benefit students' language learning. However, an achievement gap is still present among students. Given that vocabulary skills have been shown to predict literacy outcomes in the primary grades (Dickinson & Porche, 2011; Snow et al., 2007), the critical importance of supporting oral language teaching and learning in elementary grades cannot be overstated.

The overarching research question of this thesis aims at examining ways to support language learning in language-minority learners in the early elementary years. This topic is theoretically situated within a social-interactionist framework and methodologically through a researcher-practitioner partnership. Each of the three papers addresses features that can support language learning. In the following sections, the main findings are presented and discussed before conclusions and possible implications for practice are elaborated.

5.1.1 The effects of a researcher–practitioner partnership intervention

Oral language skills are the foundation for learning and participation in social contexts. A key aspect when conducting intervention research within language learning concerns whether it is possible to improve specific skills. This researcher–practitioner partnership oral language intervention (paper 1) suggested that providing language-minority learners with explicit, intense and structured instructions that help them understand basic curricular concepts through interaction can prove successful in unlocking access to the curriculum. Over the course of an eight-week program, teachers provided children multiple opportunities to participate in language-rich conversations about semantic categories and semantically related words, basic grammar and extended discourse (talk beyond the immediate here-and-now). These were all tasks chosen to facilitate language learning through the use of listening and speaking skills. The overall effect size of 0.35 is encouraging, particularly as an effect size of 0.25 or larger is considered 'substantively important' in educational research (What Works

Clearinghouse, 2014, p.14). Although a vast number of interventions have revealed positive impacts on students' language skills, these effects are primarily tied to knowledge of words taught during the intervention. Transfer to distal skills (standardized measures) rarely occurs (Elleman et al., 2009; Marulis & Neuman, 2010). This disparity may be explained by children's need to practice newly acquired skills before results can be identified on distal language outcomes, but it may also reflect the complexity of the intervention program (narrow versus broad). The present findings are in line with the recent meta-analysis by Rogde et al. (2019), where small effects were found on standardized measures of vocabulary and language comprehension with studies scoring high on quality implementation showing a larger effect size. Possible explanations for the overall effect in the present research study may be related to the group size, dosage, and quality of the implementation. First, children learn language through interaction with more knowledgeable others (Grøver et al., 2019; Halliday, 1993; Hoff, 2006). Given the flexible organizational framework of L2 instruction in Norwegian elementary schools (brief lessons in small groups or an assistant in the classroom), group size can make a difference in students' language learning. The intervention groups ranged from four to six students, and gave children numerous opportunities to use their language compared to the usual case in a classroom setting. L2 instruction in small groups also allows teachers to be responsive and tune in to students' contributions and needs to a larger extent. In comparison, Rogde et al. (2019) found significant differences in favour of small groups when comparing group size across language interventions. Second, the amount and quality of linguistic input play a pivotal role in students' language learning (Hoff, 2006; Rowe, 2012; Snow, 2014). Dosage might explain the effect given the intensity and the duration of the intervention program (64 lessons of 30-minute lessons over eight consecutive weeks). Students who are acquiring an L2 need continuous exposure. Offering students daily lessons, sometimes two times a day, may have provided the immersion needed to understand basic features of their L2. The emphasis on exposure aligns with previous research on L2 instruction (August et al., 2018; Carlo et al., 2004; Silverman, 2007). Finally, teachers adhered to the intervention components ($M = 94.5$ %). These findings comply with the Rogde et al.'s (2019) meta-analysis.

Furthermore, the overall finding from this research study is important considering the minimal amount of time spent on oral language instruction in classrooms (Nelson et al., 2015; Wright & Neuman, 2014), and how time spent can affect pace and improve learning. Proficiency in the language of instruction is critical for academic achievement (Nagy &

Townsend, 2012). A prerequisite for acquiring proficiency is a solid foundation of everyday language skills, particularly sophisticated words and extended discourse skills (Dickinson & Tabors, 2001; Uccelli et al., 2018). Teachers' awareness and knowledge of the discrepancy between everyday language skills and academic language skills are essential regarding the present achievement gap among students. School environments rely on explicit instruction to bridge students' oral language skills to the levels needed for comprehension of oral and written texts. The malleability of academic language further supports the importance of explicit instruction to level the differences in proficiency among students (Uccelli et al., 2015). This support includes helping children use academic language orally and through reading and writing. Thus, there is a need to move beyond vocabulary to improve students' skills (Proctor et al., 2019; Uccelli et al., 2015).

5.1.2 Modifications in teachers' instructional talk

Information about the quality of interaction and the type of linguistic input students are exposed to is essential in interventions targeting language learning. For language-minority learners, teachers may in many cases be the main source of L2 input. Given the lack of explicit language instruction in school (Biemiller, 2006; Nelson et al., 2015; Wright & Neuman, 2014), the value of conducting qualitative analysis of what went on during intervention lessons to improve instructional practices is supported. Moreover, the intervention program in this research study incorporated intervention components for teachers to implement that differed in the extent to which the interventions were scripted: scripted (labeling and repeated exposures of pictures of words), softly scripted (definitions and relationships among words), and minimally scripted (extended discourse; see paper 2). By examining teacher–students exchanges in the beginning and the end of the intervention period, characteristics of teacher instructional talk were identified and related to variations in scriptedness. Results revealed modifications of teacher instructional talk in relation to the softly and minimally scripted components of the program, whereas the scripted components did not change. As teachers are more likely to simplify their talk with language-minority learners (Aarts et al., 2016), this is an important finding regarding ways to support language learning. Including components that vary in scriptedness can encourage teachers to adapt their instructional talk based on students' progress throughout the program. These components may prevent teachers' tendencies to deviate from scripts during implementation of intervention programs, and in turn lead to better instructional quality. Kim et al. (2017) is one of the few

studies that encouraged teacher adaptations through guidance within specific intervention components. Their findings also revealed that students in the adapted condition of the program performed better than those receiving the core condition of the program. Thus, teachers should, to some extent, be invited to make decisions about changes based on their knowledge of students' skills and mastery, if these changes are in accordance with the theoretical framework and components driving the intervention. For instance, if students master one of the main components after a short amount of training, teachers should be able to adapt the instruction to the next level within this domain. Otherwise, repeated use of the same component with minimal variation, will most likely result in children getting bored and may affect their motivation to learn. Researcher–practitioner partnerships, such as the present study, integrate ways of adapting instruction through iterative processes between researchers and practitioners. Thus, these iterative processes unpack the competency, awareness of skills, and activities needed to support language learning by bridging knowledge from practitioners and researchers. This type of collaboration results in strategies that meet teachers' needs which may reduce possible deviations from the program during implementation.

5.1.3 Quality of assessment in education

Assessment has a key role in education as it examines what a learner knows and can do which further informs instruction. Assessment is a systematic process of arriving at an understanding of children's skills and knowledge. This research study used assessment in multiple ways: screening students for eligibility and establishing baseline skills, singling out components for use in the intervention program, examining effects of the program, and assessing the quality of an assessment tool. All of which can be helpful in finding ways to support students' language learning in school.

Language assessment tools utility depends on not only quality but also the assessor's knowledge and experience with the phenomenon examined. This examination and the resulting evidence can guide educational policies for assessment and instruction. Quality reflects the assessment tool's consistency and stability, and the extent to which the results may be valid for measuring language skills. Lack of knowledge or experience with assessment may increase the risk of under- and over-identification of students, and result in children missing out on much needed support. Moreover, examining and reporting criteria for assessment quality, unfortunately have not been a common practice for several assessment tools used in Norwegian elementary schools (Arnesen et al., 2019). For instance, paper 3

revealed that NSL can be used as a language assessment tool for monolingual and language-minority learners in grades one through four based on an analysis of the factor structure, reliability, construct validity and criterion validity. This finding is important given the previous tendency of assessment tools to be standardized on monolingual samples who may not be representative of today's multicultural society (i.e., language-minority learners). Furthermore, a recent systematic review of assessment usage in elementary schools revealed that information derived from informal assessments (i.e., teacher ratings) rather than formal assessments is primarily used to direct instruction (Arnesen et al., 2019). Given the limited amount of time spent on explicit language instruction in school, it is fundamental to illuminate how formal high-quality assessment tools can help inform decision-making to improve practice. This work also involves shedding light on current assessment approaches and practices frequently used within and across schools.

5.2 Limitations

Conducting research within educational contexts is not free of obstacles. Examining ways to support language learning in young language-minority learners reveals limitations in this research study that must be addressed. The main limitations relate to sample size, professional development, and comparison of the findings with students' national test scores.

5.2.1 Sample size

To become eligible for participation in the intervention (paper 1), students with native languages other than Norwegian and Sami had to score 1.5 standard deviation or lower on the language assessment tool NSL. Given the importance of sample size when examining effects in research, the recruitment goal was 200 students. However, consent was received from only 137 of the 203 students who fulfilled the eligibility criteria. This may be due to the limited time frame for parents to give their consent for participation (four weeks), but it can also reflect how parents were approached. Information letters were provided in Norwegian and 10 of the most common minority languages in the municipality (Urdu, Somali, Arabic, Kurdish, Turkish, Tamil, Vietnamese, Polish, English, and Albanian). Given that the final sample comprised 31 minority languages, several participants may have been lost due to proficiency in Norwegian and English (participants who spoke native language other than the 10 most common minority languages received information in Norwegian and English). Another explanation relates to the specific guidelines from the NSD regarding content information and

language style in the letter. These requirements make the information letter difficult to comprehend and may affect the response rate. In addition, four of the 20 schools were excluded because few students scored below the eligibility threshold. Given the sample selection criteria, the findings from the intervention are not generalizable to students scoring above the 1.5 standard deviation threshold.

Analysis of the teacher instructional talk was based on language exchanges occurring in the intervention group. Thus, the sample size (15 teachers and 65 students) was too small to estimate the effects of teachers' instructional talk on children's outcomes beyond the overall findings in paper 1. Moreover, analyzing what went on in the business as usual group could have revealed valuable comparisons of characteristics in teacher instructional talk across settings (group versus classroom, intervention versus business as usual) to support language learning. This topic should be addressed in future research.

5.2.2 Professional development

Another central feature of intervention studies relates to the amount of professional development offered (Zaslow, 2010). Teachers were offered only seven hours of professional development over one day combined with two weeks of preparation time before the implementation. During the implementation, teachers were offered the opportunity to correspond through email and over the telephone when needed. This decision was based on limited resources. Given the increasing amount of studies offering professional development as an introductory workshop followed by individual coaching throughout the intervention (Wasik & Hindman, 2011), it would have been interesting to see how variations in professional development could affect intervention effects. Furthermore, the teacher sample comprised highly experienced teachers. Comparing less experienced teachers with highly experienced teachers is important to address in future research to see whether components varying in scriptedness are significantly different across groups. Additionally, comparing larger teacher samples illuminates what and how teachers learn from professional development, and who benefits most from professional development.

5.2.3 Comparing findings with students' national test results

Given the relationship between children's language and literacy skills (Dickinson & Porche, 2011; Snow et al., 2007), it would have been preferable to include results from students' national test scores on literacy skills to compare the scores with the measures that assess

language skills in the intervention. This feature relates to the intervention study (paper 1) and the examination of the language assessment tool NSL (paper 3). Regarding the intervention, national tests could have supplemented information about students' progress from a longitudinal perspective, because data were collected throughout only one school year. Furthermore, comparing the NSL with the national tests could have provided additional information concerning the NSL's criterion validity.

5.3 Conclusions and Future Implications

Today's classrooms are composed of a highly diverse group of students, and studies have continuously documented a lack of optimal language support in school (Dickinson & Porche, 2011; Dickinson & Tabors, 2001; Gámez, 2015; Snow, 2017). The growing number of language-minority learners who have not yet acquired the level of proficiency needed to understand the language of instruction further bolster the need to intervene. Language develops through participation in communicative contexts, and language-minority learners must be immersed in language-rich environments continuously to reach proficiency. Language-rich environments can be optimized when tailored to students' background and knowledge, learning goals and the classroom context. This support should be characterized by open-ended questions (Wasik et al., 2006), exposure to sophisticated words, expansions of students' utterances, introduction of new information in the conversations as well as participation in extended discourse (Dickinson & Porche, 2011). These features can help students master the language of instruction in school.

Intervention studies situated within a researcher-practitioner partnership offer an avenue for linking research with practice in ways that may have implications for educational practice and policy for how to optimize language learning, in particular, determining what works for whom and under what condition based on knowledge gained from researchers and practitioners. This researcher-practitioner partnership study merged quantitative and qualitative methods to examine three ways in which language learning can be supported for young language-minority learners. First, the oral language intervention was effective in enhancing students' language skills. Key features focused on exposing children to language-rich environments that allowed them to practice their speaking and listening skills, in particular, improving word knowledge through categorization supplemented with grammatical and pragmatic skills. The combination of the amount and quality of linguistic input, dosage (64 30-minute lessons over eight weeks) and instruction in small groups (one

teacher and four to six students) are components that can help bolster language-minority learners' proficiency in the language of instruction.

Second, measures of implementation fidelity illuminated the extent to which the intervention was implemented as intended (high levels of structural fidelity were found) and how well teachers executed it (process fidelity examined the quality of the language exchanges). These measures of fidelity function as a control for various factors that might affect learning during the implementation process and thus the interpretation of the results. Providing children with linguistic input that is adaptable (i.e., components varying in scriptedness) based on their progress throughout the intervention program is vital. This is one advantage of using components varying in scriptedness featured in this research study. Thus, the qualitative part of the study generates a hypothesis of how differentiating scriptedness in intervention components can support language learning in young language-minority learners. Future research should examine possible impacts of components varying in scriptedness on students' language outcomes by expanding the sample size.

Third, identifying students in need of support relies on quality assessment tools. Such tools also inform researchers and practitioners about skills in need of support. The NSL is an assessment tool that was found to fulfil these requirements. Within the Norwegian context, early identification to support language-minority learner's learning is important, particularly through new benchmarks for assessment, in kindergarten and school (Ministry of Education and Research, 2019). An assessment tool that is standardized on a multicultural sample similar to the NSL can support language learning in school.

In conclusion, the language exposure and environments students interact with are fundamental for learning in school, particularly as oral language proficiency in L2 is critical for language-minority learners' literacy development (Uccelli & Páez, 2007). Considering the lack of optimal language support found in many classrooms, it seems important to find ways in which information on how children develop their language and acquire knowledge can be incorporated and used more efficiently to support language learning in practice. One way may be through partnerships based on iterative collaboration processes between researchers and practitioners. Thus, the findings from this research study are important for educational practice and policy regarding what can be done to efficiently support language-minority learners in the early elementary years.

References

Aarts, R., Demir-Vegter, S., Kurvers, J., & Henrichs, L. (2016). Academic language in shared book reading: Parent and teacher input to mono- and bilingual preschoolers. *Language Learning, 66*(2), 263–295. doi:10.1111/lang.12152

Aarts, R., Demir, S., & Vallen, T. (2011). Characteristics of academic language register occurring in caretaker-child interaction: Development and validation of a coding scheme. *Language Learning, 61*, 1173–1221. doi:10.1111/j.1467-9922.2011.00664.x

Agder, C. T., Hoyle, S. M., & Dickinson, D. K. (2004). Locating learning in in-service education for preschool teachers. *American Educational Research Journal, 41*(4), 867–900. doi:10.3102/00028312041004867

Arnesen, A., Braeken, J., Ogden, T., & Melby-Lervåg, M. (2019). Assessing children's social functioning and reading proficiency: A systematic review of the quality of educational assessment instruments used in Norwegian elementary schools. *Scandinavian Journal of Educational Research, 63*(3), 465–490. doi:10.1080/00313831.2017.1420685

August, D., Artzi, L., & Barr, C. (2016). Helping ELLs meet standards in English language arts and science: An intervention focused on academic vocabulary. *Reading and Writing Quarterly, 32*(4), 373–396. doi:10.1080/10573569.2015.1039738

August, D., Artzi, L., Barr, C., & Francis, D. (2018). The moderating influence of instructional intensity and word type on the acquisition of academic vocabulary in young English language learners. *Reading and Writing, 31*(4), 965–989. doi:10.1007/s11145-018-9821-1

August, D., Carlo, M. S., Dressler, C., & Snow, C. E. (2005). The critical role of vocabulary development for English language learners. *Learning Disabilities Research and Practice, 20*(1), 50–57. doi:10.111/j.1540-5826.2005.00120.x

Beck, I. L., & McKeown, M. G. (2007). Increasing young low-income children's oral vocabulary repertoires through rich and focused instruction. *The Elementary School Journal, 107*(3), 251–271. doi:10.1086/511706

Biemiller, A. (2005). Size and sequence in vocabulary development: Implications for choosing words for primary grade vocabulary instruction. In E. H. Hiebert & M. L.

Kamil (Eds.), *Teaching and learning vocabulary: Bringing research to practice* (pp. 227–248). Mahwah, NJ: Erlbaum.

Biemiller, A. (2006). Vocabulary development and instruction: A prerequisite for school learning. In D. K. Dickinson & S. B. Neuman (Eds.), *Handbook of early literacy research* (Vol. 2, pp. 41–51). New York, NY: Guilford Press.

Biemiller, A., & 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

Bierman, K. L., Domitrovich, C. E., Nix, R. L., Gest, S. D., Welsh, J. A., Greenberg, M. T., . . . Gill, S. (2008). Promoting academic and social-emotional school readiness: The Head Start REDI program. *Child Development, 79*(6), 1802–1817. doi:10.1111/j.1467-8624.2008.01227.x

Bleses, D., Højen, A., Dale, P. S., Piasta, S. B., Justice, L. M., Dybdal, L., . . . Haghish, E. F. (2018). Effective language and literacy instruction: Evaluating the importance of scripting and group size components. *Early Childhood Reading Quarterly, 42*(1), 256–269. doi:10.1016/j.ecresq.2017.10.002

Bowers, E. P., & Vasilyeva, M. (2011). The relation between teacher input and lexical growth of preschoolers. *Applied Psycholinguistics*, *32*, 221–241. doi:10.1017/S0142716410000354

Bowne, J. B., Yoshikawa, H., & Snow, C. E. (2016). Experimental impacts of a teacher professional development program in early childhood on explicit vocabulary instruction across the curriculum. *Early Childhood Research Quarterly, 34*, 27–39. doi:10.1016/j.ecresq.2015.08.002

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York, NY: Guilford Press.

Capin, P., Walker, M., Vaughn, S., & Wanzek, J. (2018). Examining how treatment fidelity is supported, measured, and reported in K–3 reading intervention research. *Educational Psychology Review, 30*(3), 885–919. doi:10.1007/s10648-017-9429-z

Carlisle, J. F., Kelcey, B., & Berebitsky, D. (2013). Teachers' support of students' vocabulary learning during literacy instruction in high poverty elementary schools. *American Educational Journal, 50*, 1360–1371. doi:10.3102/0002831213492844

Carlo, M. S., August, D., McLaughlin, B., Snow, C. E., Dressler, C., Lippman, D. N., . . . White, C. E. (2004). Closing the gap: Addressing the vocabulary needs of English-

language learners in bilingual and mainstream classrooms. *Reading Research Quarterly*, *39*(2), 188–215. doi:10.158/RRQ.39.2.3

Cash, A. H.., Cabell, S. Q., Hamre, B. K., Decoster, J., & Pianta, R. C. (2015). Relating prekindergarten teacher beliefs and knowledge to children's language and literacy development. *Teaching and Teacher Education, 48,* 97–105. doi:10.1016/j.tat.2015.02.003

Coburn, C. E., Penuel, W. R., & Geil, K. E. (2013). *Research-practice partnerships: A strategy for leveraging research for educational improvement in school districts*. New York, NY: William T. Grant Foundation. Retrieved from http://learndbir.org/resources/Coburn-Penuel-Geil-2013.pdf

Coyne, M. D., McCoach, D. B., Loftus, S., Zipoli, Jr., R., & Kapp, S. (2009). Direct vocabulary instruction in kindergarten: Teaching breadth versus depth. *Elementary School Journal, 110*(1), 1–18. doi:10.1086/598840

Coyne, M. D., McCoach, D. B., Loftus, S., Zipoli, Jr., R., Ruby, M., Crevecoeur, Y. C., & Kapp, S. (2010). Extended vocabulary instruction in kindergarten: Investigation transfer effects. *Journal of Research on Educational Effectiveness, 3*(2), 93–120. doi:10.1080/19345741003592410

Coyne, M. D., McCoach, D. B., Ware, S., Austin, C. R., Loftus-Rattan, S. M., & Baker, D. L. (2019). Racing against the vocabulary gap: Matthew effects in early vocabulary instruction and intervention. *Exceptional Children, 85*(2), 163–179. doi:10.1177/0014402918789162

Cummins, J. (1984). Language proficiency and academic achievement revisited: A response. In C. Rivera (Ed.), *Language proficiency and academic achievement* (pp. 71–76). Clevedon, England: Multilingual Matters.

Cummins, J. (2001). *Negotiation identities: Education for empowerment in a diverse society* (2nd ed.). Los Angeles: California Association for Bilingual Education.

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*, 23–45. doi:10.1016/S0272-7358(97)00043-3

Darling-Hammond, L., Banks, J., Zumwalt, K., Gomez, L., Sherin, M. G., Griesdorn, J., & Finn, L.-E. (2005). Educational goals and purposes: Developing a curricular vision for teaching. In L. Darling-Hammond & J. Bransford (Eds.), *Preparing teachers for a*

changing world. What teachers should learn and be able to do (pp. 169–200). San Francisco, CA: Jossey-Bass.

Dickinson, D. K., & Caswell, L. (2007). Building support for language and early literacy in preschool classrooms through in-service professional development: Effects of the Literacy Environment Enrichment Program (LEEP). *Early Childhood Research Quarterly, 2*(2), 243–260. doi:10.1016/j.ecresq.2007.03.001

Dickinson, D. K., Collins, M. F., Nesbitt, K., Toub, S. S., Hassinger-Das, B., Hadley, E. B., . . . Golinkoff, R. B. (2019). Effects of teacher-delivered book reading and play on vocabulary learning and self-regulation among low-income preschool children. *Journal of Cognition and Development, 20*(2), 136–164. doi:10.1081/15248372.2018.1483373

Dickinson, D. K., Hofer, K. G., Barnes, E. M., & Grifenhagen, J. F. (2014). Examining teachers' language in Head Start classrooms from a systematic linguistics approach. *Early Childhood Research Quarterly, 29*(3), 231–244. doi:10.1016/j.ecresq.2014.02.006

Dickinson, D. K., & Porche, M. V. (2011). Relation between language experiences in preschool classrooms and children's kindergarten and fourth-grade language and reading abilities. *Child Development*, *82*(3), 870–886. doi:10.1111/j.1467-8624.2011.01576.x

Dickinson, D. K., & Tabors, P. O. (2001). *Beginning literacy with language: Young children learning at home and school* (pp. 1–25). Baltimore, MD: Brookes.

Donovan, M. S., & Snow, C. (2018). Sustaining research-practice partnerships: Benefits and challenges of a long-term research and development agenda. In B. Bevan & W. R. Penuel (Eds.), *Connecting research and practice for educational improvement: Ethical and equitable approaches* (pp. 33–50). New York, NY: Routledge.

Durlak, J. A. (2010). The importance of doing well in whatever you do: A commentary on the special section, "Implementation research in early childhood education." *Early Childhood Research Quarterly, 25*(3), 348–357. doi:10.1016/j.ecresq.2010.03.003

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation of program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*, 327–350. doi:10.1007/s10464-008-9165-0

50

Elleman, A. M., Lindo, E. J., Morphy, P., & Compton, D. L. (2009). The impact of vocabulary instruction on passage-level comprehension of school-age children: A meta-analysis. *Journal of Research on Educational Effectiveness*, *2*(1), 1–44. doi.10.1080/19345740802539200

Fricke, S., Burgoyne, K., Bowyer-Crane, C., Kyriacou, M., Zosimidou, A., Maxwell, L., . . . Hulme, C. (2017). The efficacy of early language intervention in mainstream school settings: A randomized controlled trial. *Journal of Child Psychology and Psychiatry, 58*(10), 1141–1151. doi:10.1111/jcpp.12737

Fricke, S., Bowyer-Crane, C., Haley, A. J., Hulme, C., & Snowling, M. J. (2013). Efficacy of language intervention in the early years. *Journal of Child Psychology and Psychiatry*, *54*(3), 280–290. doi:10.1111/jcpp.12010

Frøyen, W., Ahmadinia, S., Heller, M. C., & Skjåk, O. (2011). *Norsk som læringsspråk, NSL* [Norwegian as a language for learning]. Oslo, Norway: PPT Oslo, Utdanningsetaten.

Frøyen, W., Ahmadinia, S., Heller, M. C., Skjåk, O., & Namvar, N. (2015). *Norsk som læringsspråk redigert, NSL-R* [Norwegian as a language for learning revised, NSL-R]. Oslo, Norway: PPT Oslo, Utdanningsetaten.

Gámez, P. (2015). Classroom-based English exposure and English language learners' expressive language skills. *Early Childhood Research Quarterly, 31*(2), 135–146. doi:10.1016/j.ecresq.2015.01.007

Gámez, P., & Lesaux, N. K. (2012). The relation between exposure to sophisticated and complex language and early-adolescent English-only and language minority learners' vocabulary. *Child Development, 83*(4), 1316–1331. foi:10.1111/j.1467-8624.2012.01776.x

Gersten, R., Dimino, J., Jayanthi, M., Kim, J. S., & Santoro, L. E. (2010). Teacher study groups: Impact of the professional development model on reading instruction and student outcomes in first grade classrooms. *American Educational Research Journal, 47*(3), 694–739. doi:10.3102/0002831209361208

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children, 71*(2), 149–164. doi:10.1177/001440290507100202

Grøver Aukrust, V. (2007). Young children acquiring second language vocabulary in preschool group-time: Does amount, diversity, and discourse complexity of talk

matter? *Journal of Research in Childhood Education, 22*(1), 17–37. doi:10.1080/02568540709594610

Grøver, V., Rydland, V., Gustafsson, J.-E., & Snow, C. E. (2020). Shared book reading in preschool supports bilingual children's second-language learning: A cluster-randomized trial. *Child Development*. doi:10.1111/cdev.13348

Grøver, V., Uccelli, P., Rowe, M. L., & Lieven, E. (2019). Learning through language. In V. Grøver, P. Uccelli, M. L. Rowe, & E. Lieven (Eds.), *Learning through language: Towards an educationally informed theory of language learning* (pp. 1–15). Cambridge, England: Cambridge University Press.

Gutiérrez, K. D., & Penuel, W. R. (2014). Relevance to practice as a criterion for rigor. *Educational Researcher, 43*(1), 19–23. doi:10.3102/0013189X13520289

Hadley, E. B., Dickinson, D. K., Hirsh-Pasek, K., & Golinkoff, R. M. (2018). Building semantic networks: The impact of a vocabulary intervention on preschoolers' depth of word knowledge. *Reading Research Quarterly, 54*(1), 41–61. doi:10.1002/rrq.225

Hagen, Å. M., Melby-Lervåg, M., & Lervåg, A. (2017). Improving language comprehension in preschool children with language difficulties: A cluster randomized trial. *Journal of Child Psychology and Psychiatry, 58*(10), 1132–1140. doi:10.1111/jcpp.12762

Halliday, M. A. K. (1975). *Learning how to mean: Explorations in the development of language*. London, England: Arnold.

Halliday, M. A. K. (1993). Towards a language-based theory of learning. *Linguistics and Education, 5*(2)*,* 93–116. doi:10.1016/0898-5898(93)90026-7

Hamre, B. K., Justice, L. M., Planta, R. C., Kilday, C., Sweeney, B., Downer, J. T., & Leach, A. (2010). Implementation fidelity of MyTeachingPartner literacy and language activities: Association with preschoolers' language and literacy growth. *Early Childhood Research Quarterly, 25*(3), 329–347. doi:10.1016/j.ecresq.2009.07.002

Harn, B., Parisi, D., & Stoolmiller, M. (2013). Balancing fidelity with flexibility and fit: What do we really know about fidelity of implementation in school? *Exceptional Children, 79*(2), 181–193. doi:10.1177/001440291307900204

Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday life of young American children*. Baltimore, MD: Brookes.

Hassinger-Das, B., Ridge, K., Parker, A., Golinkoff, R. M., Hirsh-Pasek, K., & Dickinson, D. K. (2016). Building vocabulary knowledge in preschoolers through shared book

reading and gameplay. *Mind, Brain and Education, 10*(2), 71–80.
https://doi.org/10.1111/mbe.12103

Hill, H. C., & Erickson, A. (2019). Using implementation fidelity to aid in interpreting
program impacts: A brief review. *Educational Researcher, 48*(9), 590–598.
doi:10.3102/0013189X19891436

Hoff, E. (2003). The specificity of environmental influences: socioeconomic status affects
early vocabulary development via maternal speech. *Child Development, 74,* 1368–
1378.

Hoff, E. (2006). How social contexts support and shape language development.
Developmental Review, *26*(1), 55–88. doi:10.1016/j.dr.2005.11.002

Hoff, E. (2014). *Language development* (5th ed.). Australia: Wadsworth Cengage Learning.

Hulme, C., & Snowling, M. J. (2009). *Developmental disorders of language learning and
cognition*. Oxford, England: Wiley-Blackwell.

Huttenlocher, J., Vasilyeva, M., Cymerman, E., & Levine, S. (2002). Language input and
child syntax. *Cognitive Psychology, 45*(3), 337–374. doi:10.1016/S0010-
0285(02)00500-5

Jackson, B., Larzelere, R., Clair, L., Corr, M., Fichter, C., & Egertson, H. (2006). The impact
of HeadsUp! Reading on early childhood educators' literacy practices and preschool
children's literacy skills. *Early Childhood Research Quarterly, 21*(2), 213–226.
doi:10.1016/j.eresq.2006.04.005

Jayanthi, M., Dimino, J., Gersten, R., Taylor, M. J., Haymond, K., Smolkowski, K., &
Newman-Gonchar, R. (2018). The impact of teacher study groups in vocabulary on
teaching practice, teacher knowledge, and student vocabulary knowledge: A large-
scale replication study. *Journal of Research in Educational Effectiveness, 11*(1), 83–
108. doi:10.1080/19345747.2017.1327625

Justice, L. M., Mashburn, A. J., Hamre, B. K., & Pianta, R. C. (2008). Quality of language
and literacy instruction in preschool classrooms serving at-risk pupils. *Early
Childhood Research Quarterly, 23*(1), 51–68. doi:10.1016/j.ecresq.2207.09.004

Kim, J. S., Burkhauser, M. A., Quinn, D. M., Guryan, J., Kingston, H. C., & Aleman, K.
(2017). Effectiveness of structured teacher adaptations to an evidence-based summer
literacy program. *Reading Research Quarterly, 52*(4), 443–467. doi:10.1002/rrq.178

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). New
York, NY: The Guilford Press.

Landry, S. H., Anthony, J. L., Swank, P. R., & Monseque-Bailey, P. (2009). Effectiveness of comprehensive professional development for teachers of at-risk preschoolers. *Journal of Educational Psychology, 101*(2), 448–465. doi:10.1037/a0013842

Lawrence, J. F., Crosson, A. C., Paré-Blagoev, E. J., & Snow, C. E. (2015). Word generation randomized trial: Discussion mediates the impact of program treatment on academic word learning. *American Educational Research Journal, 52*(4), 750–786. doi:10.3102/0002831215579485

Leseman, P. P. M., Scheele, A. F., Mayo, A. Y., & Messer, M. H. (2007). Home literacy as a special language environment to prepare children for school. *Zeitschrift für Erziehungswissenschaft, 10*, 334–355. doi:10.1007/s11618-007-0040-9

Lonigan, C. J., Farver, J. M., Phillips, B. N., & Clancy-Menchetti, J. (2011). Promoting the development of preschool children's emergent literacy skills: A randomized evaluation of a literacy-focused curriculum and two professional development models. *Reading and Writing, 24*, 305–337. doi:10.1007/s11145-009-9214-6

Mancilla-Martinez, J., & Lesaux, N. K. (2011). The gap between Spanish speakers' word reading and word knowledge: A longitudinal study. *Child Development, 82*(5), 1544–1560. doi:10.1111/j.1467-8624.2011.01633.x

Markussen-Brown, J., Juhl, C. B., Piasta, S. B., Bleses, D., Højen, A., & Justice, L. M. (2018). The effects of language- and literacy-focused professional development on early educators and children: A best-evidence meta-analysis. *Early Childhood Research Quarterly, 38*(1), 97–115. doi:10.1016/j.eresq.2016.07.002

Marulis, L. M., & 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. doi:10.3102/0034654310377087

Marulis, L. M., & Neuman, S. B. (2013). How vocabulary interventions affect young children at risk: A meta-analytic review. *Journal of Research on Educational Effectiveness, 6*(3), 223–262. doi:10.1080/19345747.2012.755591

McKeown, M. G., & Beck, I. 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.2014.06002

Mendive, S., Weiland, C., Yoshikawa, H., & Snow, C. E. (2016). Opening the black box: Intervention fidelity in a randomized trial of a preschool teacher professional

development program. *Journal of Educational Psychology, 108*(1), 130–145. doi:10.1037/edu0000047

Messick, S. (1995). Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741–749. doi:10.1037/0003-066x.50.9.741

Milburn, T. F., Girolametto, L., Weitzman, E., & Greenberg, J. (2014). Enhancing preschool educators' ability to facilitate conversations during shared book reading. *Journal of Early Childhood Literacy, 14*(1), 105–140. doi:10.1177/1468798413478261

Ministry of Education and Research. (2019). Tett på – tidlig innsats og inkluderende fellesskap i barnehage, skole og SFO. Meld. St. 6 (2019-2020) [White paper 6 (2019-2020)]. Oslo, Norway: Departementets servicesenter.

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

Mowbray, C., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation, 24*, 315–340. doi:10.1016/S1098-2140(03)00057-2

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.)*.* Los Angeles, CA: Muthén & Muthén.

Nagy, W. E., & Scott, J. A. (2000). Vocabulary processes. In M. L. Kamil, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 269–284). New York, NY: Routledge.

Nagy, W., & Townsend, D. (2012). Words as tools: Learning academic vocabulary as language acquisition. *Reading Research Quarterly, 47*(1), 91–108. doi:10.1002/RRQ.011

National Committee for Research Ethics in the Social Sciences and the Humanities, Norway. (2006). *Guidelines for research ethics in the social science law and the humanities*. Oslo, Norway: NESH.

Nelson, K. L., Dole, J. A., Hosp, J. L., & Hosp, M. K. (2015). Vocabulary instruction in K-3 low-income classrooms during a reading reform project. *Reading Psychology, 36*(2), 145–172. doi:10.1080/02702711.2013.839485

Neugebauer, S., Coyne, M., McCoach, B., & Ware, S. (2017). Teaching beyond the intervention: The contribution of teacher language extensions to vocabulary learning

in urban kindergarten classrooms. *Reading & Writing, 30,* 543–567. doi:10.1007/s11145-016-9689-x

Neuman, S. B. (2011). The challenge of teaching vocabulary in early education. In S. B. Neuman, & D. K. Dickinson (Eds.), *Handbook of early literacy research* (Vol. 3, pp. 358–372). New York, NY: The Guilford Press.

Neuman, S. B., & Cunningham, L. (2009). The impact of professional development and coaching on early language and literacy instructional practices. *American Educational Research Journal, 46*(2), 532–566. doi:10.3102/0002831208328088

Neuman, S. B., & Dwyer, J. (2011). Developing vocabulary and conceptual knowledge for low-income preschoolers: A design experiment. *Journal of Literacy Research, 43*(2), 103–129. doi:10.1177/1086296X11403089

Neuman, S. B., Newman, E. H., & Dwyer, J. (2011). Educational effects of vocabulary intervention on preschoolers' word knowledge and conceptual development: A cluster-randomized trial. *Reading Research Quarterly*, *46*(3), 249–272. doi:10.1598/RRQ.46.3.3

Neuman, S. B., Pinkham, A., & Kaefer, T. (2015). Supporting vocabulary teaching and learning in prekindergarten: The role of educative curriculum materials. *Early Education and Development*, *26*(7), 988–1011. doi:10.1080/10409289.2015.1004517

Nielsen, D. C., & Friesen, L. D. (2012). A study of the effectiveness if a small-group intervention on the vocabulary and narrative development of at-risk kindergarten children. *Reading Psychology, 33*(3), 269–299. doi:10.1080/02702711.2010.508671

Odom, S. L., Fleming, K., Diamond, K., Lieber, J., Hanson, M., Butera, G., … Marquis, J. (2010). Examining different forms of implementation and in early childhood curriculum research. *Early Childhood Research Quarterly, 25*(3), 314–328. doi:10.1016/j.ecresq.2010.03.001

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research, 78*, 33–84. doi:10.3102/0034654307313793

Phillips, B. M., Ingrole, S. A., Burris, P. W., & Tabulda, G. (2017). Investigating predictors of fidelity of implementation for a preschool vocabulary and language curriculum. *Early Child Development and Care, 187*(3-4), 542–553. doi:10.1080/03004430.2016.1251428

Phillips Galloway, E., & Lesaux, N. (2017). A matter of opportunity: Language reading development during early childhood for dual-language learners. In N. Kucirkova, C. E. Snow, V. Grøver, & C. McBride (Eds.), *The Routledge international handbook of early literacy education. A contemporary guide to literacy teaching and interventions in a global context* (pp. 26–39). Oxon, NY: Routledge.

Piasta, S. B., Connor, C. M., Fishman, B. J., & Morrison, F. J. (2009). Teacher's knowledge of literacy concepts, classroom practices, and student reading growth. *Scientific Studies of Reading, 13*(3), 224–248. doi:10.1080/1088430902851364

Piasta, S., Justice, L., McGinty, A., Mashburn, A., & Slocum, L. (2015). A comprehensive examination of preschool teachers' implementation fidelity when using a supplemental language and literacy curriculum. *Child & Youth Care Forum, 44*(5), 731-755. doi:10.1007/s10566-015-9305-2

Pollard-Durodola, S., Gonzalez, J., Simmons, D., Kwok, O., Taylor, A., Davis, M., . . . Simmons, L. (2011). The effects of an intensive shared book-reading intervention for preschool children at risk for vocabulary delay. *Exceptional Children, 77*(2), 161–183. doi:10.1177/001440291107700202

Power, T. J., Blom-Hoffman, J., Clarke, A. T., Riley-Tillman, T. C., Kelleher, C., & Manz, P. H. (2005). Reconceptualizing intervention integrity: A partnership-based framework of linking research with practice. *Psychology in the Schools, 42*(5), 495–507. doi:10.1002/pits.20087

Proctor, C. P., Silverman, R. D., Harring, J. R., Jones, R. L., & Hartranft, A. M. (2019). Teaching bilingual learners: Effects of a language-based reading intervention on academic language and reading comprehension in grades 4 and 5. *Reading Research Quarterly, 55*(1), 95–122. doi:10.1002/rrq258

Putnik, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review, 41*, 71–90. doi:10.1016/j.dr.2016.06.004

Rogde, K., Hagen, Å. M., Melby-Lervåg, M., & Lervåg, L. (2019). The effect of language comprehension training on standardized tests: A systematic review. *Campbell Systematic Reviews.*

Rogde, K., Melby-Lervåg, M., & Lervåg, A. (2016). Improving the general language skills of second-language learners in kindergarten: A randomized controlled trial. *Journal of*

    *Research on Educational Effectiveness, 9(*1), 150–170.
    doi:10.1080/19345747.2016.1171935

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development*, *83*(5), 1762–1774. doi:10.1111/j.1467-8624.2012.01805.x

Rydland, V., Grøver, V., & Lawrence, J. F. (2014). The second-language trajectories of Turkish immigrant children in Norway from ages five to ten: The role of preschool talk exposure, maternal education, and co-ethnic concentration in the neighbourhood. *Journal of Child Language, 41*(2), 352–381. doi:10.1017/S0305000912000712

Sattler, J. M. (2018). *Assessment of children: Cognitive foundations and applications*. La Mesa, CA: Author.

Schachter, R. E. (2017). Early childhood teachers' pedagogical reasoning about how children learn during language and literacy instruction. *International Journal of Early Childhood, 49*(1), 95–111. doi:10.1007/s13158-017-0179-3

Schachter, R., Spear, C. F., Piasta, S., Justice, L. M., & Logan, J. A. R. (2016). Early childhood educator's knowledge, and children's language- and literacy-learning opportunities: What is the connection? *Early Childhood Research Quarterly, 36*, 281–294. doi:10.1016/j.eresq.2016.01.008

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Belmont, CA: Wadsworth Cengage Learning.

Silverman, R. D. (2007). Vocabulary development of English-language and English-only learners in kindergarten. *The Elementary School Journal, 107*(4), 365–383. doi:10.1086/516669

Silverman, R. D., & Crandell, J. D. (2010). Vocabulary practices in pre-kindergarten and kindergarten classrooms. *Reading Research Quarterly, 45*(3), 318–340. doi:10.1598/RRQ.45.3.3

Silverman, R. D., & Hines, S. (2009). The effects of multimedia enhanced instruction on the vocabulary of English language learners and non-English language learners in pre-kindergarten through second grade. *Journal of Educational Psychology, 101*, 305–314. doi:10.1037/a0014217

Silverman, R. D., Martin-Beltran, M., Peercy, M. M., Hartranft, A. M., McNeish, D. M.,

Artzi, L., & Nunn, S. (2017). Effects of a cross-age peer learning program on the vocabulary and comprehension of English learners and non-English learners in elementary school. *The Elementary School Journal, 117*(3), 485–512. doi:10.1086/690210

Silverman, R. D., Proctor, C. P., Harring, J. R., Doyle, B., Mitchell, M. A., & Meyer, A. G. (2014). Teachers' instruction and student's vocabulary and comprehension: An exploratory study with English monolingual and Spanish-English bilingual students in grades 3–5. *Reading Research Quarterly, 49*(1), 31–60. doi:10.1002/rrq.63

Snow, C. E. (1990). The development of definitional skill. *Journal of Child Language, 17*, 697–710.

Snow, C. E. (2014). Extended discourse in first and second language acquisition: A challenge and an opportunity. *Journal of Japanese Linguistics, 30*(1), 7–14. doi:10.1515/jjl-2014-0102

Snow, C. E. (2015). 2014 Wallace Foundation Distinguished Lecture: Rigor and realism: Doing educational science in the real world. *Educational Researcher, 44*(9), 460–466. doi:10.3101/0013189X15619166

Snow, C. E. (2017). The role of vocabulary versus knowledge in children's language learning: A fifty-year perspective. *Infancia y Aprendizaje, 40*(1), 1–18. doi:10.1080/02103702.2016.1263449

Snow, C. E., Lawrence, J. F., & White, C. (2009). Generating word knowledge of academic language among urban middle school students. *Journal of Research on Educational Effectiveness, 2*(4), 325–344. doi:10.1080/19345740903167042

Snow, C. E., Porche, M. V., Tabors, P. O., & Harris, S. R. (2007). *Is literacy enough? Pathways to academic success for adolescents.* Baltimore, MD: Brookes.

Snow, C. E., Tabors, P. O., & Dickinson, D. K. (2001). Language development in the preschool years. In D. K. Dickinson & P. O. Tabors (Eds.), *Beginning literacy with language: Young children learning at home and school* (pp. 1–25). Baltimore, MD: Brookes.

Snow, C. E., & Uccelli, P. (2009). The challenge of academic language. In D. R. Olson & N. Torrance (Eds.), *The Cambridge handbook of literacy* (pp. 112–133). Cambridge, England: Cambridge University Press.

Stahl, S. A., & Fairbanks, M. M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research, 56*(1), 72–110. doi:10.3102/00346543056001072

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38*(6), 934–947. Doi:10.1037/012-1649.38.6.934

Swanson, E., Wanzek, J., Haring, C., Ciullo, S., & McCulley, L. (2011). Intervention fidelity in special and general education research journals. *The Journal of Special Education, 47*(1). 3–13. doi:10.1177/0022466911419516

Tangen, R. (2014). Balancing ethics and quality in educational research- the ethical matrix method. *Scandinavian Journal of Educational Research, 58*(6), 678–694. doi:10.1080/00313831.2013.821089

Thorndike, R. M., & Thorndike-Christ, T. (2014). *Measurement and evaluation in psychology and education* (8th ed.). Harlow, England: Pearson.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Townsend, D., Filippini, A., Collins, P., & Biancarosa, G. (2012). Evidence of the importance of academic word knowledge for the academic achievement of diverse middle school students. *The Elementary School Journal, 112*(3), 497–518. doi:10.1086/663301

Troyer, M. (2019). Teachers' adaptations to and orientation towards an adolescent literacy curriculum. *Journal of Curriculum Studies, 51*(2), 202–228. doi:10.1080/00220272.2017.1407458

Truckenmiller, A. J., Park., J., Dabo, A., & Wu, Y.-C.N. (2019). Academic language instruction for students in grade 4 through 8: A literature synthesis. *Journal of Research on Educational Effectiveness, 12*(1), 135–159. doi:10.1080/19345747.2018.1536773

Uccelli, P., Demir-Lira, Ö. E., Rowe, M., Levine, S., & Goldin-Meadow, S. (2018). Children's early decontextualized talk predicts academic language proficiency in mid-adolescence. *Child Development, 90*(5), 1650–1663. doi:10.1111/cdev.13034

Uccelli, P., & Paéz, M. M. (2007). Narrative and vocabulary development of bilingual

children from kindergarten to first grade: Developmental changes and associations among English and Spanish skills. *Language, Speech, and Hearing Services in School, 38*(3), 225–236. doi:0161-1461/07/3803-0225

Uccelli, P., Phillips Galloway, E., Barr, C. D., Meneses, A., & Dobbs, C. L. (2015). Beyond vocabulary: Exploring cross-disciplinary academic-language proficiency and its association with reading comprehension. *Reading Research Quarterly, 50*(3), 337–356. doi:10.1002/rrq.104

Vygotsky, L. S. (1986). *Thought and language*. Cambridge, MA: MIT Press.

Wasik, B. A., Bond, M. A., & Hindman, A. (2006). The effects of a language and literacy intervention on Head Start children and teachers. *Journal of Educational Psychology, 98*(1), 63–74. doi:10.1037/0022-0663.98.1.63

Wasik, B. A., & Hindman, A. H. (2011). Improving vocabulary and pre-literacy skills of at-risk preschoolers through teacher professional development. *Journal of Educational Psychology, 103*(2), 455–469. doi:10.1037/a0023067

Wasik, B. A., & Hindman, A. H. (2018). Increasing preschoolers' vocabulary development through a streamlined teacher professional development intervention. *Early Childhood Research Quarterly*. doi:10.1016/j.ecresq.2018.11.001

Weizman, Z., & Snow, C. E. (2001). Lexical input as related to children's vocabulary acquisition: Effects of sophisticated exposure and support for meaning. *Developmental Psychology, 37,* 265–279.

What Works Clearinghouse. (2014). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, What Works Clearinghouse. Retrieved from https://ies.ed.gov/ncee/wwc/.

Wright, T. S., & Neuman, S. B. (2014). Paucity and disparity in kindergarten oral vocabulary instruction. *Journal of Literacy Research, 46*(3), 330–357. doi:10.1177/1086296X14551474

Zaslow, M., Tout, K., Halle, T., Whittaker, J. V., Lavelle, B., & Trends, C. (2010). *Toward the identification of features of effective professional development of early childhood educators. Literature review*. U.S. Department of Education. Office of Planning, Evaluation and Policy Development/Policy and Program Studies Service. Retrieved from https://www2.ed.gov/rschstat/eval/professional-development/literature-review.pdf

Appendices

Approval from the Norwegian Social Science Data Services (NSD)

Mia Cecilie Heller
Institutt for pedagogikk Universitetet i Oslo
Postboks 1092 Blindern 0317 OSLO

Vår dato: 02.06.2016   Vår ref: 48601 / 3 / AMS   Deres dato: Deres ref:

TILBAKEMELDING PÅ MELDING OM BEHANDLING AV PERSONOPPLYSNINGER

Vi viser til melding om behandling av personopplysninger, mottatt 09.05.2016. Meldingen
gjelder prosjektet: 48601 Andrespråkslæring i skolens begynneropplæring: en
intervensjonsstudie
Behandlingsansvarlig: Universitetet i Oslo, ved institusjonens øverste leder
Daglig ansvarlig: Mia Cecilie Heller

Personvernombudet har vurdert prosjektet, og finner at behandlingen av personopplysninger
vil være regulert av § 7-27 i personopplysningsforskriften. Personvernombudet tilrår at
prosjektet gjennomføres.

Personvernombudets tilråding forutsetter at prosjektet gjennomføres i tråd med opplysningene
gitt i meldeskjemaet, korrespondanse med ombudet, ombudets kommentarer samt
personopplysningsloven og helseregisterloven med forskrifter. Behandlingen av
personopplysninger kan settes i gang.

Det gjøres oppmerksom på at det skal gis ny melding dersom behandlingen endres i forhold
til de opplysninger som ligger til grunn for personvernombudets vurdering.
Endringsmeldinger gis via et eget skjema,
http://www.nsd.uib.no/personvern/meldeplikt/skjema.html. Det skal også gis melding etter tre
år dersom prosjektet fortsatt pågår. Meldinger skal skje skriftlig til ombudet.

Personvernombudet har lagt ut opplysninger om prosjektet i en offentlig database,
http://pvo.nsd.no/prosjekt.
Personvernombudet vil ved prosjektets avslutning, 31.01.2020, rette en henvendelse angående
status for behandlingen av personopplysninger.

Vennlig hilsen
Kjersti Haugstvedt   Anne-Mette Somby

Kontaktperson: Anne-Mette Somby tlf: 55 58 24 10
Vedlegg: Prosjektvurdering

Personvernombudet for forskning

Prosjektvurdering - Kommentar
Prosjektnr: 48601

FORMÅL
Prosjektet har som hensikt å øke kunnskapen om hvordan skolen kan støtte barns (med norsk som andrespråk) språklæring. Målet med prosjektet er å vurdere et spesifikt språkkurs effekt på elevenes språkferdigheter.

INFORMASJON OG SAMTYKKE
Utvalget (foreldre og lærere) informeres skriftlig om prosjektet og samtykker til deltakelse. Informasjonsskrivet er godt utformet.
Merk at når barn skal delta aktivt, er deltagelsen alltid frivillig for barnet, selv om de foresatte samtykker. Barnet bør få alderstilpasset informasjon om prosjektet, og det må sørges for at de forstår at deltakelse er frivillig og at de når som helst kan trekke seg dersom de ønsker det.

SENSITIVE OPPLYSNINGER
Det er ikke krysset av for at det kan framkomme opplysninger om etnisk bakgrunn i meldeskjemaet. Personvernombudet har tatt høyde for at dette kan framkomme, og har endret dette punktet i meldeskjemaet.

INFORMASJONSSIKKERHET
Personvernombudet legger til grunn at forsker etterfølger Universitetet i Oslo sine interne rutiner for datasikkerhet.

PROSJEKTSLUTT OG ANONYMISERING
Forventet prosjektslutt er 31.01.2020. Ifølge prosjektmeldingen skal innsamlede opplysninger da anonymiseres. Anonymisering innebærer å bearbeide datamaterialet slik at ingen enkeltpersoner kan gjenkjennes. Det gjøres ved å: - slette direkte personopplysninger (som navn/koblingsnøkkel) - slette/omskrive indirekte personopplysninger (identifiserende sammenstilling av bakgrunnsopplysninger som f.eks. bosted/arbeidssted, alder og kjønn) - slette digitale lyd-/bilde- og videoopptak

Forespørsel om deltakelse i forskningsprosjektet:

Andrespråkslæring i skolens begynneropplæring: en intervensjonsstudie

«Ønsker du å bidra med kunnskap om språkutvikling hos barn med norsk som andrespråk?»

Til rektor

Bakgrunn og formål med studien
Vi tar kontakt for å informere om forskningsprosjektet «Andrespråkslæring i skolens begynneropplæring: en intervensjonsstudie», og for å spørre om skolen vil delta i prosjektet. Prosjektet er finansiert av Universitetet i Oslo og gjennomføres ved Institutt for pedagogikk.

Studien gjennomføres på 1. og 2.trinn ved en rekke skoler i Oslo. Hensikten med studien er å frembringe ytterligere kunnskap om hvordan skolen kan støtte barns språklæring. Gjennom studien håper vi å bidra til kunnskap som vil være viktig for skolene når de skal legge til rette for språkopplæring for barn med norsk som andrespråk, ved at det gjennomføres en effektvurdering av Norsk intensivt språkkurs (NISK). Et godt språk gir elevene et grunnleggende utgangspunkt for tilegnelse av lese- og skriveferdigheter på skolen.

Osloskolen har siden høsten 2014 tatt i bruk språkkurset NISK. I samarbeid med Utdanningsetaten, vil skoler som tilbyr NISK-kurs plukkes ut til å delta i denne studien. Din skole, er trukket ut til å delta i studien, og vi ønsker derfor å invitere til et samarbeid hvor vi vil følge 1. og 2.trinnselevs språklige utvikling gjennom et skoleår. Forespørselen om deltakelse i studien sendes til skoler i Oslo som gir 1. og 2.trinnselever tilbud om NISK-kurs.

Hva innebærer deltakelse i studien?
Deltakelse i studien innebærer at skolen:
- vurderer hvilke elever som skal motta NISK på bakgrunn av resultatene fra kartleggingsverktøyet Norsk som læringsspråk (NSL). Elevene må skåre innenfor kategorien risiko eller vanske på NSL.
- formidler informasjon om studien til foresatte ved bruk av ranselpost
- bistår med innsamling av utfylte samtykkeerklæringer og spørreskjemaer fra foresatte
- gjennomfører NISK-kurs for elever hvor foresatte har samtykket til deltakelse i forskningsstudien (kurs for elever som ikke har samtykke vil eventuelt foregå parallelt og i en annen NISK-gruppe)
- følger PPTs kriterier for kartlegging av elevene før og etter NISK med NSL (NB! Lærer som har NISK-kurs kan ikke gjennomføre kartlegging med NSL)
- følger PPTs kriterier for gjennomføring av undervisningsøktene (64 undervisningsøkter innenfor nivåene A, B, C og D over 8 uker, samt underveiskartlegging med NISK-skjemaer for hvert av nivåene)

- NISK-lærer (lærer som er ansvarlig for kurset og har mottatt opplæring) gjennomfører lydopptak av seks NISK-økter (diktafon vil bli tildelt hver skole) som forskningsassistentene samler inn. Disse opptakene vil bli benyttet for å vurdere innholdet i kurset
- NISK-lærer besvarer et spørreskjema som omhandler rapportering knyttet til innholdet i NISK-øktene
- stiller et grupperom eller liknende til disposisjon for forskningsassistentene til supplerende kartlegging av NISK-elevene som har samtykke fra foresatte til å delta i studien. Denne kartleggingen består av 4 ulike oppgaver som alle vil gi viktig informasjon om elevenes språkferdigheter. Kartlegging av eleven vil ta ca 30 minutter hver gang

Det vil bli gjennomført tre kartlegginger av elevene som deltar i studien, en før intervensjonsgruppen mottar NISK, og en etter at NISK er gjennomført. I tillegg vil det blir gjennomført en kartlegging i april-juni, etter at kontrollgruppen også har mottatt NISK. Dette innebærer at skolen kartlegger elevene med NSL 3 ganger i løpet av studien.

Hva skjer med informasjonen om skolen?
Alle personopplysninger vil bli behandlet konfidensielt. Det betyr blant annet at vi har strenge krav til hvordan vi lagrer materialet vi samler inn. Bare de som er direkte tilknyttet forskningsprosjektet vil se materialet, og skolens, læreres og elevenes navn lagres ikke sammen med materialet. Det vil derfor ikke være mulig å kjenne igjen enkeltbarn, lærere eller skoler når resultatene fra undersøkelsene presenteres som vitenskapelige artikler og som en doktoravhandling.

Forskningsprosjektet skal etter planen avsluttes i 31.01. 20. Listen med deltakernavn og lydopptakene vi samler inn slettes når forskningsprosjektet avsluttes.

Frivillig deltakelse
Det er frivillig å delta i studien. Dersom dere har spørsmål til studien, ta kontakt med Mia C. Heller på tlf. 23 85 55 17 eller e-post: m.c.heller@iped.uio.no

Studien er meldt til Personvernombudet for forskning, Norsk samfunnsvitenskapelig datatjeneste AS.

Med vennlig hilsen

_____   _____
Vibeke Grøver   Mia C. Heller
Professor   PhD-stipendiat
                                                   22 85 55 17

Samtykkeerklæring

Jeg har lest informasjonsbrevet og gir med dette tillatelse til at
_____ skole kan delta i undersøkelsen om språkutvikling:

Dato:

_____
Rektors signatur

På forhånd takk for hjelpen!

Forespørsel om deltakelse i forskningsprosjektet:

Andrespråkslæring i skolens begynneropplæring: en intervensjonsstudie

«Ønsker du å bidra med kunnskap om språkutvikling hos barn med norsk som andrespråk?»

Til foresatte med barn på 1. og 2.trinn med norsk som andrespråk

Vi tar kontakt for å informere om forskningsprosjektet «Andrespråkslæring i skolens begynneropplæring: en intervensjonsstudie», og for å spørre om dere vil at barnet deres skal delta i prosjektet. Prosjektet er finansiert av Universitetet i Oslo og gjennomføres ved Institutt for pedagogikk.

Bakgrunn og formål med studien
Studien gjennomføres på 1. og 2.trinn ved en rekke skoler i Oslo. Hensikten med studien er å frembringe mer kunnskap om hvordan skolen kan støtte barns språklæring. Gjennom studien håper vi å bidra til kunnskap som vil være viktig for skolene når de skal legge til rette for språkopplæring for barn med norsk som andrespråk. Et godt språk gir et grunnleggende utgangspunkt for tilegnelse av lese- og skriveferdigheter på skolen.

Osloskolen har siden høsten 2014 tatt i bruk språkkurset «Norsk intensivt språkkurs» (NISK). Dette er et undervisningstilbud som gis til barn med norsk som andrespråk. I samarbeid med Utdanningsetaten, har skoler som tilbyr NISK-kurs blitt plukket ut til å delta i denne studien.

Ditt barns skole, er trukket ut til å delta i studien, og vi ønsker derfor å invitere til et samarbeid hvor vi vil følge ditt barns språklige utvikling gjennom et skoleår. Forespørselen om deltakelse i studien sendes til alle foresatte med barn som får særskilt norskopplæring på skolen.

Hva innebærer deltakelse i studien?

Dersom du samtykker til deltakelse, innebærer det at ditt barns språk kartlegges før og etter NISK-kurset, totalt tre ganger. Språkkartleggingen tar ca 30 minutter hver gang, den foregår på skolen og gjennomføres av forskningsassistenter. Barn pleier å like denne type oppgaver. Vi vil også ta et lydopptak av barnet i denne kartleggingen, i tillegg til at det vil bli gjennomført lydopptak i undervisningen med læreren. Hensikten med lydopptakene er å få informasjon om barnets talespråk.

Barna som deltar i studien deles tilfeldig inn i to grupper, gruppe 1 og gruppe 2. En av gruppene får NISK-kurs først. Gruppe 2 får NISK-kurs når gruppe 1 har gjennomført NISK-kurset. Undervisningen foregår på skolen og vil bli gitt av lærere.

For å få ytterligere kunnskap om barnets språkbruk, ber vi dere om å fylle ut vedlagte spørreskjema om språkbruk, familie og interesser.

Hva skjer med informasjonen om barnet ditt?

Alle personopplysninger vil bli behandlet konfidensielt. Det betyr blant annet at vi har strenge krav til hvordan vi lagrer materialet vi samler inn. Bare de som er direkte tilknyttet forskningsprosjektet vil se materialet, og barnets navn lagres ikke sammen med materialet. Det vil derfor ikke være mulig å kjenne igjen enkeltbarn eller skoler når resultatene fra studien presenteres som vitenskapelige artikler og som en doktoravhandling.

Forskningsprosjektet skal etter planen avsluttes i 30.01.20. Listen med deltakernavn og lydopptakene vi samler inn slettes når forskningsprosjektet avsluttes.

Frivillig deltakelse

Det er frivillig å delta i studien, og dere kan når som helst trekke dere uten å oppgi noen grunn ved en e-post eller telefon til Mia C. Heller. Dersom dere trekker dere, vil barnets navn slettes fra listen over deltakere og materialet som gjelder ditt barn slettes. Det vil ikke ha noen innvirkning på forholdet til skolen om dere velger å ikke delta eller om dere velger å trekke dere på et senere tidspunkt. Dere har rett til innsyn i opplysninger som registreres om eget barn. Dersom dere ikke ønsker å delta i studien, vil barnet deres likevel få tilbud om NISK-kurs på skolen. Kurset vil da ikke være en del av dette forskningsprosjektet.

Dersom dere har spørsmål til studien, ta kontakt med Mia C. Heller på tlf. 23 85 55 17 eller e-post: m.c.heller@iped.uio.no

Studien er meldt til Personvernombudet for forskning, Norsk samfunnsvitenskapelig datatjeneste AS.

Med vennlig hilsen

_____   _____

Vibeke Grøver   Mia C. Heller
Professor   PhD-stipendiat
                                                             22 85 55 17

Samtykkeerklæring

Jeg har lest informasjonsbrevet og gir med dette tillatelse til at mitt barn kan delta i undersøkelsen om språkutvikling:

_____   _____
   _____
Barnets navn   Født: Kjønn

_____
Morsmål

_____   _____
Skole   Klasse

Dato:

_____
Foresattes signatur

Forespørsel om deltakelse i forskningsprosjektet:

Andrespråkslæring i skolens begynneropplæring: en intervensjonsstudie

«Ønsker du å bidra med kunnskap om språkutvikling hos barn med norsk som andrespråk?»

Til lærer

Bakgrunn og formål med studien
Vi tar kontakt for å informere om forskningsprosjektet «Andrespråkslæring i skolens begynneropplæring: en intervensjonsstudie», og for å spørre om skolen vil delta i prosjektet. Prosjektet er finansiert av Universitetet i Oslo og gjennomføres ved Institutt for pedagogikk.

Studien gjennomføres på 1. og 2.trinn ved en rekke skoler i Oslo. Hensikten er å bidra til kunnskap om hvordan skolen kan legge til rette for språkopplæring for barn med norsk som andrespråk, ved å gjennomføre en effektvurdering av språkkurset Norsk intensivt språkkurs (NISK). Et godt språk gir elevene et grunnleggende utgangspunkt for tilegnelse av lese- og skriveferdigheter på skolen.

Osloskolen har siden høsten 2014 tatt i bruk NISK. I samarbeid med Utdanningsetaten, vil skoler som tilbyr NISK-kurs plukkes ut til å delta i studien. Din skole er trukket ut til å delta i studien, og vi ønsker derfor å invitere til et samarbeid hvor vi vil følge 1. og 2.trinnselevers språklige utvikling gjennom et skoleår. Forespørselen om deltakelse i studien sendes til skoler i Oslo som gir 1. og 2.trinnselever tilbud om NISK-kurs.

Hva innebærer deltakelse i studien?
Deltakelse i studien innebærer at du:
- vurderer hvilke elever som skal motta NISK på bakgrunn av resultatene fra kartleggingsverktøyet Norsk som læringsspråk (NSL). Elevene må skåre innenfor kategorien risiko eller vanske på NSL.
- formidler informasjon om studien til foresatte ved bruk av ranselpost
- bistår med innsamling av utfylte samtykkeerklæringer og spørreskjemaer fra foresatte
- gjennomfører NISK-kurs for elever hvor foresatte har samtykket til deltakelse i forskningsstudien (kurs for elever som ikke har samtykke vil eventuelt foregå parallelt og i en annen NISK-gruppe)
- følger PPTs kriterier for kartlegging av elevene før og etter NISK med NSL (NB! Lærer som har NISK-kurs kan ikke gjennomføre kartlegging med NSL)
- følger PPTs kriterier for gjennomføring av undervisningsøktene (64 undervisningsøkter innenfor nivåene A, B, C og D over 8 uker, samt underveiskartlegging med NISK-skjemaer for hvert av nivåene).
- NISK-lærer (lærer som er ansvarlig for kurset og har mottatt opplæring) gjennomfører lydopptak av seks NISK-økter (diktafon vil bli tildelt hver skole) som

forskningsassistentene samler inn. Disse opptakene vil bli benyttet for å vurdere innholdet i kurset

- NISK-lærer besvarer et/to korte spørreskjemaer; et som omhandler rapportering knyttet til innholdet i NISK-øktene og et som omhandler vurdering av PPTs NISK-kursing av lærerne

Hva skjer med informasjonen om deg?

Alle personopplysninger vil bli behandlet konfidensielt. Det betyr blant annet at vi har strenge krav til hvordan vi lagrer materialet vi samler inn. Bare de som er direkte tilknyttet forskningsprosjektet vil se materialet, og skolens, lærernes og elevenes navn lagres ikke sammen med materialet. Det vil derfor ikke være mulig å kjenne igjen enkeltbarn, lærere eller skoler når resultatene fra undersøkelsene presenteres som vitenskapelige artikler og som en doktoravhandling.

Forskningsprosjektet skal etter planen avsluttes i 31.01. 20. Listen med deltakernavn og lydopptakene vi samler inn slettes når forskningsprosjektet avsluttes.

Frivillig deltakelse

Det er frivillig å delta i studien. Dersom dere har spørsmål til studien, ta kontakt med Mia C. Heller på tlf. 23 85 55 17 eller e-post: m.c.heller@iped.uio.no

Studien er meldt til Personvernombudet for forskning, Norsk samfunnsvitenskapelig datatjeneste AS.

Med vennlig hilsen

_____   _____
Vibeke Grøver   Mia C. Heller
Professor   PhD-stipendiat
                                                    22 85 55 17

Samtykkeerklæring

Jeg har lest informasjonsbrevet om studien, og er villig til å delta.

Dato:

_____
Signatur

På forhånd takk for hjelpen!

SPØRRESKJEMA FORESATTE

Takk for at dere deltar i denne undersøkelsen om språkutvikling hos barn med norsk som andrespråk. Dersom dere ikke ønsker å besvare et av spørsmålene, gå videre til neste spørsmål.

1. Hvem har fylt ut spørreskjemaet?

    Mor

    Far

    Begge

Fortell oss litt om familien:

2. Hva er barnets navn?_____

3. Hvilken skole går barnet på? _____   klasse:_____

4. Er barnet født i Norge?

    Ja

    Nei

Dersom barnet er født i et annet land, hvor mange år har barnet bodd i Norge?_____ år

5. Har barnet søsken?3

    Ja

    Nei

6. Har barnet gått i barnehage?

    Ja

    Nei

Hvis ja, hvor gammelt var barnet da det begynte i barnehagen?_____ år _____ måneder.

7. Fikk barnet tilbud om morsmålsstøtte i barnehagen?

    Ja

    Nei

8. Er mor født i Norge?

    Ja

    Nei

Dersom født i et annet land, hvor lenge har mor bodd i Norge? _____ år.

9. Er far født i Norge?
>   Ja
>   Nei

Dersom født i et annet land, hvor lenge har far bodd i Norge? _____ år.

10. Foresattes utdanning:

a) Mor:
>   1-7 år (Barneskole)
>   8-10 år (Ungdomsskole)
>   11-13 år (Videregående skole)
>   Bachelorgrad (høyskole/universitet)
>   Mastergrad eller mer (høyskole/universitet)

b) Far:
>   1-7 år (Barneskole)
>   8-10 år (Ungdomsskole)
>   11-13 år (Videregående skole)
>   Bachelorgrad (høyskole/universitet)
>   Mastergrad eller mer (høyskole/universitet)

Fortell oss litt om språkbruk i familien

11. Hva er barnets morsmål? _____

12. Hvilket språk snakker mor med barnet?
>   mest morsmål
>   både morsmål og norsk
>   mest norsk

13. Hvilket språk snakker far med barnet?
>   mest morsmål
>   både morsmål og norsk
>   mest norsk

14. Hvilket språk snakker barnet til mor?
>   mest morsmål
>   både morsmål og norsk
>   mest norsk

15. Hvilket språk snakker barnet til far?

    mest morsmål

    både morsmål og norsk

    mest norsk

16. Hvilket språk snakker barnet med søsken?

    mest morsmål

    både morsmål og norsk

    mest norsk

17. Snakkes det andre språk i hjemmet?

    Ja

    Nei

    Hvis ja, hvilket? _____

18. Får barnet morsmålsopplæring utenfor skolen og hjemmet?

    Ja

    Nei

Fortell oss litt om barnets interesser

19. Sett kryss på <u>tre</u> av aktivitetene nedenfor som barnet liker å gjøre.

    Se på tv/filmer

    Spille spill (på data, mobil, X-box, Playstation etc.)

    Tegne

    Å bli lest for

    Lese/se i bøker på egenhånd

    Utelek

    Puslespill

    Innelek

    Leke med ball

    Sykle

    Sang

    Konstruksjonslek (Bygge/lage ting med klosser)

    Høre på musikk

    Danse

    Annet, spesifiser: _____

20. Er det annen informasjon du ønsker å gi oss om barnet ditt?

Tusen takk for utfyllingen av spørreskjemaet. Skjemaet skal returneres til skolen i vedlagte konvolutt.

Ta kontakt med Mia C. Heller dersom det er spørsmål: m.c.heller@iped.uio.no eller tlf: 22 85 55 17

PART II

The Papers

Paper 1:   Heller, M.C., Lervåg, A., & Grøver, V. (2019). Oral language intervention in Norwegian schools serving young language-minority learners: A randomized trial. *Reading Research Quarterly*, 54(4), 531-552. doi:10.1002/rrq.248

Paper 2:   Heller, M.C., & Grøver, V. (2019). Teachers' Instructional Talk in a Partly Scripted Language Intervention Targeting Young Language-Minority Learners: Developments over time. Manuscript submitted for publication.

Paper 3:   Heller, M.C., & Lervåg, A. (2020). Kartlegging av språkferdigheter for elever på 1.-4.trinn: En vurdering av kartleggingsverktøyet Norsk Som Læringsspråk (NSL). ["Assessing students' oral language skills in grades 1-4: An evaluation of the assessment tool Norwegian as a Language for Learning (NSL)"].

II

Teachers' Instructional Talk in a Partly Scripted Language Intervention Targeting Young

Language-Minority Learners: Developments Over Time

Running head: TEACHERS' INSTRUCTIONAL TALK

Abstract

Despite huge investments in interventions designed to support oral language skills in early childhood and beyond, many of the interventions fail to identify impacts on children's language learning. Programs may have limited impact because they do not sufficiently succeed in supporting teachers' instructional talk, and thus, more efficiently promote children's language learning. The present study examined the extent to which 15 teachers in Norway implementing a language intervention program designed to enhance students' second-language learning demonstrated changes in their instructional talk over the 8-week program. The program consisted of scripted parts (labeling pictures of targeted words, repeated exposures), as well as soft scripted parts (word relations and definitions) and minimally scripted parts (narratives and explanations that extended the here-and-now). Teachers received professional development that qualified them to implement the program. Analysis of modifications in teachers' instructional talk was based on audio-recorded small-group interactions, comparing characteristics of teacher talk at the beginning and end of the 8-week program. Results revealed that teachers' instructional talk developed to include more word definitions and extended discourse, talk categories aligned with the less scripted parts of the intervention. Conversely, teacher talk during the scripted parts of the program did not change.

Keywords: instructional talk, scripted components, oral language learning, interventions

Teachers' Instructional Talk in a Partly Scripted Language Intervention Targeting Young

Language-Minority Learners: Developments Over Time

From a social-interactionist theoretical perspective, language learning is embedded within and results from the complexity of oral interactions that children participate in during and outside school (Author & Colleagues, 2019; Halliday, 1993; Hoff, 2006; Ninio & Snow, 1996). A basic quality of language intervention programs within the social-interactionist approach is the teacher–student language-promoting interactions the programs support (Dickinson & Porche, 2011; Justice, Jiang, & Strasser, 2018; Michener, Proctor, & Silverman, 2018). In modern multilingual societies, identifying qualities of language intervention programs that support teacher capacity to engage minority learners in language-promoting interaction is crucial.

Features of interactions that support language learning have been adapted into intervention programs by encouraging conversations using strategies such as open-ended questions (Wasik, Bond, & Hindman, 2006; Zucker, Justice, Piasta, & Kaderavek, 2010), reinforcement and repetitions (Carlo et al., 2004; Roberts & Neal, 2004), and expansions of children's utterances and the use of sophisticated vocabulary words (Dickinson & Porche, 2011). However, studies examining the efficacy of language intervention programs on children's oral language skills have reached inconsistent findings. Wasik and Hindman (2011) found that an intervention program based on teacher professional development improved the quality of teachers' instructional talk (e.g., modeling language or providing children opportunities to use language), and that variation in the quality of teacher talk was associated with children's language gains. Meta-analyses by Elleman, Lindo, Morphy, and Compton (2009) and Marulis and Neuman (2010) revealed positive impacts of language interventions on some oral language measures but not on others. Other language intervention studies did not find effects on child outcomes despite huge investments (e.g., Yoshikawa et al., 2015).

Language Learning and Scripted Intervention Components

Key decisions in the construction of language intervention programs for young learners concern the detailing of teacher manuals. Some intervention programs offer detailed and manualized guidelines for how teachers should interact with students during activities; that is, the programs are strongly scripted, while other intervention programs are less detailed when it comes to how teachers' instructions are supported. We use the term *scripting* in this article to refer to prescribed, read-aloud components, which are highly structured with distinct steps and instructions to complete. Scripted versions usually contain step-by-step instruction for teachers to follow from A to B (e.g., reading out each instruction directly from the manual), and do not open up for flexible use and extensions of the targeted activities and instructions. An example is the work by Rogde, Melby-Lervåg, and Lervåg (2016), who provided teachers with detailed, word-by-word lesson scripts to follow throughout the intervention. Other intervention programs demonstrate a softer type of scripting, for example, by providing some intervention components with specified instructions for teachers to follow (i.e., scripted word definitions), while also inviting teacher adaptations that are not specifically outlined regarding what teachers should emphasize (e.g., Neugebauer, Coyne, McCoach, & Ware, 2017). Even less guidelines for instructional talk is offered in intervention programs that mostly provide teachers with examples of activities to select from within the targeted domain (Wasik & Hindman, 2018). The teacher support in these latter programs may include information on general language-enhancing strategies that teachers receive as part of professional development during workshops before the intervention, and that teachers are encouraged to use in a flexible way throughout the implementation of the intervention. For some intervention programs, a mixture of scripted and less scripted components is included (Kim et al., 2017).

Although studies have revealed that teachers can implement a program with high levels of adherence when it comes to delivering the number of lessons and targeted components in the intervention as intended, high implementation fidelity does not necessarily reflect quality instruction (Harn, Parisi, & Stoolmiller, 2013; Justice, Mashburn, Hamre, & Pianta, 2008). Less focus may be placed on the dynamic interactions that occur between the teacher and the students, and on the information provided to students (Bowne, Yoshikawa, & Snow, 2016).

We identified three studies that shed light on the impact of intervention scriptedness on teachers' instructional talk. First, Kim et al. (2017) examined the effectiveness of two conditions of the program READS, an original version with scripted components and an adapted version in which teachers were offered guidance on how to modify specific parts of the program. Findings showed that teachers in both conditions implemented the core elements with high levels of fidelity, while teachers in the adapted condition also afforded extensions and modifications in line with the program's aim. Students participating in the adapted version performed better on reading comprehension than their peers in the original condition, and were also provided with more exposure to core components of the program (e.g., longer duration or extensions of activities).

Second, a recent study compared the quality of instructional talk by examining scripted components and teachers' extensions of these components in relation to students' language growth (Neugebauer et al., 2017). Quality of instructional talk was examined on two levels: scripted instructions of definitions and word relations, and extensions of these definitions and word relations (which were not scripted but in line with the program's aim). Findings revealed that adapted instructional extensions were most beneficial in enhancing students' language growth.

Third, a large-scale Danish study compared three intervention conditions that differed in their level of scripting;: small-group and large-group conditions with softly scripted lesson plans linked to language and literacy activities, and a small-group condition that differed from the two others by being only minimally scripted. The latter intervention condition provided the same scope and sequence of instructions as for the other two conditions, but without a scripted plan to follow (Bleses et al., 2018). The impact was larger in the minimally scripted condition compared to the other two conditions. Furthermore, the minimally scripted version contributed to gains in children's vocabulary, with larger gains achieved by second-language learners compared to their monolingual peers.

Interestingly, Dickinson (2011) suggested that interventions may not have the desired impact because they fail to change the ways teachers talk with children, and thus, the teachers' capacity to support children's language learning. Relevant in that regard is Smolkowski and Gunn's (2012) work that showed teachers' ways of providing opportunities for classroom learning were stable over time. Teachers' instructional talk and ways of interacting with students reflect their cultural backgrounds and beliefs, and several studies have demonstrated that such talk may not easily be altered (Dickinson & Caswell, 2007; Dickinson & Porche, 2011; Hindman & Wasik, 2012; Neuman & Cunningham, 2009).

However, studies of whether teachers who receive professional development and participate in language intervention programs are able to modify their instructional talk over time as they deliver the program to their students, may not have distinguished sufficiently between developments in instructional talk that are associated with more scripted parts of the intervention program and developments in instructional talk that result from, and reflect, less scripted intervention components. Examining teacher instructional talk aligned with components that differ in degree of program scriptedness may help to identify qualities of

teachers' language-promoting interaction with students, and thus, also provide a basis for

future intervention programs designed to support student language learning.

The Present Study

The aim of the present study was to examine characteristics of the instructional talk of

teachers who participated in a language intervention program addressing young language-

minority students, and to assess how the teachers' instructional talk developed from the early

to the later parts of the intervention program. The teachers participated in an intervention

program that consisted of scripted and less scripted components. Various terms have been

used in the literature to describe degrees of scripting, either of an entire intervention program

or of specific components of a program, such as overly and highly scripted (Parks & Bridges-

Rhoads, 2012), soft-scripted (Bleses et al., 2018), or largely unscripted (Neugebauer et al.,

2017). Given the diversity of terms, and reflecting the characteristics of the intervention

program examined, we distinguished between three levels of program component scriptedness:

scripted components, softly scripted components, and minimally scripted components.

The study was part of a randomized controlled 8-week trial targeting oral language

learning in young language-minority learners in the early elementary school years. Children

who received the intervention showed statistical significant improvements in oral language

skills compared to children who were enrolled in the waiting-list control group, with an

overall effect size of 0.35 immediately following the intervention (Authors et al., 2019). The

intervention program included scripted components (various types of labeling with detailed

guidelines for implementation), more softly scripted components (e.g., invitations to define a

word in ways that were not prescribed or detailed in the guidelines to the teachers), and

minimally scripted components (e.g., teacher invitations to children to extend the here-and-

now, reflecting specific children's interest and experience). We expected that teachers may

use more of the less scripted components as they got to know the students in addition to

students developing their second-language skills, even within the short time span of eight weeks. If this is the case, the design of intervention studies should take into consideration that teacher talk may develop during an intervention. We examined how the components of the intervention program, varying in degrees of scriptedness, appeared in the teachers' interactions with students, and whether we could identify a development in the teachers' uptake of the various intervention components. More specifically, we asked the following two research questions:

RQ1. What characterized teachers' instructional talk in the early and late parts of the intervention?

RQ2. Did teachers modify their instructional talk during the intervention?

Methods

Participants

The participants were 15 (13 female) teachers and 65 (37 female) students. All teachers worked as resource teachers in the children's schools, but were not their classroom teachers, and therefore, did not know the students before the intervention study. Resource teachers were mainly responsible for second-language support and adaptive education in the schools. In Norway, a four-year college or university degree is required to qualify as a teacher. The participating teachers were experienced, with 11.79 years ($SD = 10.57$) of mean teaching experience in elementary schools. A total of 16 teachers participated in the randomized, controlled study (Authors et al., 2019), but one teacher did not complete the requested audio recordings on which the present study was based, and thus, was not included.

Participating students were all second-language learners receiving the intervention. To be included in the study, the children scored in the lower range (1.5 SD below the mean or lower) on a standardized language test in Norwegian, *Norsk Som Læringsspråk* (Norwegian as a Language for Learning, NSL; Colleagues & Author, 2011). The students had a mean age

of 77.05 months (SD = 6.730 months). Although the children spoke various first languages, Urdu, Arabic, Somali, Kurdish, Turkish, and Tamil accounted for 66.1% (for a more detailed description of the total sample, see Authors et al., 2019). The 15 groups had a mean of 4.40 students (SD = 0.63).

Intervention

The main part of the program had clearly scripted guidelines for implementation (the labeling of the visual material), while other program components were presented as examples of language-supporting instructional talk (softly scripted), and others represented encouragement to extend talk beyond the here-and-now (minimally scripted). The lesson design aimed at enhancing oral language learning and affording students with a baseline of words for accessing the curriculum. As all students had low levels of second-language proficiency, words that were crucial for comprehending and learning curricular topics across subjects were chosen (i.e., acquire everyday content words already known by their monolingual peers). Explicit explanations of word meaning supplemented with visual material targeting semantic categories, such as "home" or "hospital," and words useful for talking about them, such as for "home," apartment, elevator, and hall, and for "hospital," wound, patient, and ambulance, were the main content in each lesson. This usage of semantic categories aligns with studies documenting the efficacy of teaching word knowledge through conceptual categories, such as interventions targeting taxonomic usage (sorting words in hierarchical order; Neuman, Newman, & Dwyer, 2011), and those examining the effects of taxonomic and thematic groups (Hadley, Dickinson, Hirsh-Pasek, & Golinkoff, 2018; Pinkham, Kaefer, & Neuman, 2014). Children received 64 30-min lessons over 8 consecutive weeks.

The professional development the teachers received comprised a one-day workshop (7 h) in which the rationale and the aim for the intervention was presented. The introduction of

the program's components was combined with practice in small groups supervised by school psychologists. In the latter, activities that were part of the scripted components of the intervention were first demonstrated, followed by the teachers practicing the skills with supervision. These activities were then discussed in plenary to clarify and provide further examples if needed. Teachers were encouraged to draw students' attention to word relations and definitions in a less prescribed form (the softly scripted components of the program), as well as to invite children to tell about events in and outside the classroom (the minimally scripted components). Finally, teachers were told to adapt their instructional talk based on students' progress throughout the intervention, and various examples were illustrated and discussed (i.e., how to extend their language use and level of abstraction). Support beyond this workshop was provided by telephone or email.

Procedures

Data collection. Teachers audiotaped their instruction and interaction with students during three lessons in the second (time point 1; t1) and seventh weeks of the intervention (time point 2; t2). We selected for transcription and analysis the first out of the three lessons in each respective category, as this was the lesson when teachers introduced for the first time the new semantic category to be worked with ("home" in week 2 and "hospital" in week 7). The approximate first 20 min of each lesson were selected for transcription (t1; $M = 18.13$ min, $SD = 3.45$; t2; $M = 17.34$ min, $SD = 1.96$). This main part of each lesson consisted of the same framework in which the targeted category and its related words were introduced for the first time through labeling (scripted) accompanied by conversations extending the here-and-now children's interest and experience (minimally scripted), in addition to talk about word relations and definitions (softly scripted).

The audio recordings were transcribed following the format of Codes for the Human Analysis of Transcripts (CHAT) from the Child Language Data Exchange System (CHILDES)

(MacWhinney, 2000). Utterances were identified by intonation, pauses, and conversational turns that indicated a break in the flow of speech. Transcriptions were limited to utterances in Norwegian. We were not able to identify with sufficient reliability individual student utterances in the audio recordings, and therefore, the transcripts did not distinguish between individual student speakers.

Coding scheme development and measures. The applied coding scheme builds on recent studies in which features of teachers' instructional talk were examined. We sought to incorporate talk categories that captured a continuum of language skills, ranging from basic word level skills at one end to extended discourse in the other, describing the variety of linguistic input students may be exposed to in classrooms.

Talk categories developed to identify how intervention components varying in scriptedness appeared in the classroom.

Talk categories aligned with scripted components of the intervention. To identify instructional talk that reflected the scripted parts of the intervention, we developed two categories that were derived and adapted from Bowne et al.'s (2016) conceptual information category (concrete information about words): simple labeling and elaborated labeling.

Simple labeling. This category was used for labeling pictures, either isolated by basic naming of the targeted word (*sykepleier*, "nurse") or through basic sentences (*dette er en sykepleier*, "this is a nurse").

Elaborated labeling. This category included information about the meaning of a word, as provided by describing illustrations, acting out the word, facts, examples, and information of what was not characteristic of the targeted word (e.g., *dette er noe barn er redd for (sprøyte),* "this is something children are afraid of (a syringe)").

A code was given for each utterance that offered either simple or elaborated labeling. The coding also applied to utterances that repeated, reinforced, or reformulated a previous utterance coded as labeling.

Talk categories aligned with the softly scripted components of the intervention. To address instructional talk reflecting the softly scripted parts of the intervention, we developed two categories, building on definition and word relation categories from studies by Bowne et al. (2016), who distinguished between definitions and examples of word usage, and Neugebauer et al. (2017), who also examined definitions and word relations. These categories, conceptual definitions and usage-based examples of definitions, differed from the labeling categories previously described in the completeness of information provided in the teacher guidelines.

Conceptual definitions. This category encompassed information about word meaning provided through developmentally relevant definitions, such as *sykepleier det er hun som passer på de syke, gir dem mat og medisiner*/ ("a nurse is someone who takes care of people that are sick, gives them food and medicine"). Teachers were not offered prescribed guidelines for conceptual definitions, and therefore, provided students with the teachers' own adapted versions. When teachers spontaneously offered synonyms for targeted words, they were also included in the conceptual definitions category, such as in the following definition of a doctor: *en lege er det samme som en doktor* ("a doctor is the same as a physician").

Usage-based examples of definitions. We applied this category to identify context-related demonstrations of word use such as in the following examples: *når man reiser til noen land så må man ta en reisevaksine (sprøyte)* ("sometimes when you travel to another country you need to get a vaccine") or *man tar også røntgenbilde av tennene hos tannlegen* ("we can also take an x-ray of the teeth at the dentist"). This category was distinguished from the former category by less formal descriptions to exemplify word usage.

In a similar vein as for the talk categories aligned with the scripted components of the intervention, each utterance identified as either a conceptual or usage-based example of definitions was counted. This also applied to utterances that repeated, reinforced, or reformulated an utterance identified as a definition.

Talk category aligned with the minimally scripted components of the intervention.

To identify talk that reflected the minimally scripted components of the intervention, and that invited student contributions extending the here-and-now, we developed the talk category extended discourse which included sequences of talk that were explanations, narratives, or a combination. This category built on the definition of extended discourse provided by Snow, Tabors, and Dickinson (2001): "talk that requires participants to develop understandings beyond the here and now and that requires the use of several utterances or turns to build a linguistic structure, such as in explanations, narratives, or pretend" (p. 2). Usually, this type of talk emerged from the targeted words (e.g., a teacher asking a follow-up question, "Has anybody been to a hospital?' or "What does vaccine mean? What do we use it for?"), but could also be introduced by a child who contributed a related story or introduced a word associated with the category. We marked in the transcripts where a sequence of extended discourse started and ended, and counted the number of utterances within the sequence. Given the second-language skills of the participating students, we included all utterances appearing within the segment, and utterances that were simple confirmations (e.g., utterances 6 and 9), negations, and repetitions of the previous utterance (e.g., utterance 5). The following example is an excerpt from a sequence identified as extended discourse, and was introduced by the teacher asking whether everybody had a home. The children followed up on her question by confirming and disagreeing. The teacher then stated that most people have a home:

Example 1

Extended Discourse Emerging from the Category Home

1. TEA: De fleste har det.
   (Most people have a home.) (pause)
2. TEA: Men det er kanskje noen som ikke har det.
   (But maybe there is someone who does not have one.)
3. TEA: Vi skal snakke om hjem, om ting som vi har hjemme.
   (We're going to talk about home, about things we have in a home.)
4. TEA: Noen bor i li (.) små hus, og noen bor i høye hus.
   (Some live in small houses; others live in apartment buildings.)
5. Child: Jeg bor i høyhus.
   (I live in an apartment building.)
6. TEA: Ja.
   (Yes.)
7. TEA: Og hvor bor +/.
   (And where does +/.)
8. CHI: Jeg bor også i høye hus.
   (I live in an apartment building, too.)
9. CHI: Jeg også.
   (Me too.)
10. TEA: Mm, og noen bor i et eget hus.
    (mm, and some live in their own house.)

Tokens and types of talk. Studies using tokens and types as indicators of talk quantity and talk diversity, respectively, have typically reported on large variations from one classroom to another, regarding teacher and student contributions (Author, 2007; Colleagues & Author, 2014). Thus, we decided to use tokens and types as a control measure of talk quantity and diversity when examining degrees of scripted instructions.

Tokens. Talk quantity was measured by the number of tokens (the sheer number of words), using the FREQ option in Computerized Child Language Analysis (CLAN; MacWhinney, 2000).

Types. Similarly, talk diversity was measured by calculating the number of different words (types) in each transcript. To develop a talk index that differed from the tokens index, we created a list of exclusion words against which all transcripts were screened. The list comprised numbers, vocal gestures ("mm," "oi"), personal names, and the 150 most common words according to a Norwegian word frequency list. All morphological variants of a word were counted as one word type (e.g., *jente-jenter* "girl" and "girls"). However, when words were derived from the same word root, but differed semantically, they were both included

(e.g., *mørk-mørket,* "dark" and "darkness"). All transcripts were examined visually to check

for misspellings that might have occurred during transcription. As the transcribed recordings

varied somewhat in duration, we developed a density measure for the tokens (number of

tokens per minute) and the types (number of types per minute).

Reliability. The first author and a research assistant independently double-coded one

transcript and discussed the applicability of the codes. When the two coders disagreed, the

final coding criteria were arrived at through discussion between them. The first author and the

research assistant then assessed the reliability of the coding scheme based on 20% of the

recordings (six recordings in total, pre- and postrecordings). Interrater reliability was

computed using Cohen's kappa, and had a mean average of .84 (ranging between .78 and .93),

which indicated substantial agreement (Landis & Koch, 1977).

Analytical Approach

Identifying characteristics of teachers' instructional talk and possible modifications of

teacher talk throughout the intervention were the aims of this study. To respond to RQ1, we

examined the frequency of talk categories in the teachers' instructional talk, in the early and

late parts of the intervention. To respond to RQ2, we used paired sample *t* tests to estimate

and compare modifications in instructional talk at the two time points. Although the research

questions addressed teachers, the student groups' talk developed during the intervention, and

we included examination of changes in student groups' talk as a check. Additionally,

measures of tokens and types of talk for teachers and the student groups were included as

check points when we examined the scripted and less scripted talk categories. These decisions

allowed us to comprehend teachers' discourse utterances more fully. As lesson duration

varied across groups, measures of talk exposure were calculated per minute, and no

statistically significant difference was found between the two time points ($t(14) = 1.013$, *p*

= .328). Complete sets of data on the five categories of teachers instructional talk measures of

linguistic input were retrieved for all teachers.

Results

To respond to the first research question, we present descriptive statistics on the

occurrence of simple labeling, elaborated labeling, conceptual definitions, usage-based

examples of definitions, and extended discourse produced by teachers (see Table 1). There

was substantial variability among the teachers across all measures at both time points. As

displayed in Table 1, predominant talk categories in the beginning of the intervention were

simple and elaborated labeling, both aligned with the scripted components of the intervention.

Talk categories aligned with softly scripted components (definitions and usage-based

examples of definitions) and minimally scripted components (extended discourse) appeared

much less frequently. In comparison, the less scripted components were often used by the end

of the intervention (t2).

[Table 1 approximately here.]

The second research question sought to identify whether changes in the appearance of

teachers' instructional talk had occurred between the two time points. To assess these aspects,

paired t tests were conducted. Statistically significant increases were found for conceptual

definitions ($t(14) = -.2.709, p = .017$), usage-based examples of definitions ($t(14) = -2.800, p$

$= .014$), and extended discourse ($t(14) = -4.649, p = .000$). However, neither simple labeling

($t(14) = 1.663, p = .119$) or elaborated labeling ($t(14) = 2.018, p = .063$) changed statistically

significantly, although the latter approached a statistically significant decrease. Simple

labeling and elaborated labeling were both talk categories aligned with the scripted

components of the intervention, while the talk categories demonstrating statistically significant increases were aligned with the softly scripted and minimally scripted components.

As students' language skills developed during the intervention (Authors et al., 2019), we also examined how these talk categories appeared in students' talk to check whether modifications in teacher talk reflects changes in student talk. Similar frequency patterns of talk categories were found in the students groups at t1 and t2 (see Table 2). Moreover, statistically significant positive correlations were found at both time points between teachers' and student groups' use of labeling (t1; $r = .72$, $p = .002$; t2; $r = .68$, $p = .005$) and extended discourse (t1; $r = .90$, $p = .000$; t2; $r = .84$, $p = .000$). We found positive statistically significant correlations at t2, but not at t1, between teachers' and student groups' use of conceptual definitions (t1; $r = -.34$, $p = .220$; t2; $r = .55$, $p = .034$) and usage-based examples of definitions (t1; $r = .38$, $p = .161$; t2; $r = .66$, $p = .007$). Elaborated labeling was positively statistically significantly correlated only at t1 (t1; $r = .54$, $p = .036$; t2; $r = -.05$, $p = .880$). Furthermore, examinations of changes in children's talk using paired sampled t tests revealed statistically significant increases in conceptual definitions ($t(14) = -2.487$, $p = .026$), usage-based examples of definitions ($t(14) = -2.851$, $p = .013$), and extended discourse ($t(14) = -4.375$, $p = .001$) in the student groups, in addition to elaborated labeling ($t(14) = -4.237$, $p = .001$). In contrast, the use of simple labeling ($t(14) = 2.111$, $p = .053$) demonstrated a marginally statistically significant decrease.

[Table 2 approximately here.]

Finally, we used the token density and types density by the teachers and the student groups as a check when examining and interpreting the prevalence and modifications of the scripted-aligned talk categories. Teachers produced, on average, 74.46 tokens per minute (*SD*

= 22.71 tokens), and 6.54 types per minute (SD = 2.17 types) at t1. By the end of the intervention, the mean number of tokens per minute was 72.91 (SD = 13.60), and 8.08 types per minute (SD = 1.53). Students as a group produced, on average, 21.32 tokens per minute (SD = 5.78 tokens), and 2.73 types per minute (SD = 1.00 types) at baseline. By the seventh week, the students produced, on average, 31.32 tokens per minute (SD = 12.70), and 4.63 types per minute (SD = 1.70). Furthermore, teachers provided the same number of tokens at both time points ($t(14)$ = .330, p = .746), while the diversity of words used was statistically significantly higher at t2 ($t(14)$ = –3.624, p = .003). In comparison, students as a group demonstrated higher density of different word types ($t(14)$ = –5.486, p = .000) and talked more (higher token density) by the end of the intervention ($t(14)$ = –4.125, p = .001). Thus, teachers allowed themselves to be less scripted as students as a group developed into more talkative interaction partners.

Discussion

This study explored patterns of teachers' instructional talk, and whether modifications of talk occurred in relation to degrees of scripting when participating in an oral language intervention for young language-minority learners in the early elementary years. In response to RQ1, we found high usage of labeling in teachers' instructional talk in the beginning of the intervention, while this feature was less prominent at t2, with definitions and extended discourse emerging. In response to RQ2, we demonstrated that teachers modified their linguistic input during the intervention period, as statistically significant changes in instructional talk were found in softly scripted and minimally scripted instructional talk (definitions, usage-based examples of definitions, and extended discourse). Scripted instructional talk (labeling) showed stability throughout the intervention. These results are important, as some teachers tend to simplify their speech with second-language learners,

which does not provide children with language-rich environments needed for learning (Aarts, Demir-Vegter, Kurvers, & Henrichs, 2016).

Dickinson (2011) documented how difficult it is to change teachers' practices, and Smolkowski and Gunn (2012) reported in a similar vein on high stability in instruction over time. We argue that the distinction between scripted and less scripted components in interventions is important to address, as it can provide insights into the mechanisms underlying change. Moreover, different levels of scripting offer teachers various directions for enriching the linguistic input in teacher–student exchanges, which can support students' language learning. Interventions that include less scripted components in addition to prescribed parts provide teachers with the autonomy and flexibility that make it feasible to adapt to developing children's progress and level of proficiency throughout the intervention. We believe the less scripted components of the present intervention allowed teachers to incorporate more advanced talk in their exchanges with students, while also inviting students to elaborate on their knowledge related to the lesson's topic. In the literature, adaptations of instructions in interventions have recently received increased attention, emphasizing a need to consider the possibility of flexible adaptations as an active ingredient in interventions. Combining different levels of scripted instructions can also give teachers greater agency, which may enhance the quality of the implementation and students' outcomes (Kim et al., 2017). This perspective concurs with findings from Neugebauer and colleagues (2017), where adaptations in the form of extensions from the scripted instructions were the most beneficial component in promoting students' language growth. In addition, when Bleses et al. (2018) compared different language interventions, varying in their degree of scripting, the less scripted version was the most effective compared to the two more scripted versions, with second-language learners having the greatest gains. Affording teachers greater agency by inviting them to adapt the intervention to the specific students they work with may provide

teachers with more opportunities to fine-tune their interactions with their students, compared to when they are asked to follow a more scripted intervention manual.

Scripted intervention components can offer an opportunity for teachers to acquire a new skills toolkit that can be incorporated in their instruction over time. Therefore, it may lead to changes in their instructional practices, which can promote children's language learning (Neugebauer et al., 2017). Furthermore, manuals provide examples of how to initiate language-rich interactions with students. For instance, some teachers may be less comfortable adapting the instruction to students' needs during the intervention. This can be due to limited competence or experience with strategies promoting language learning. Accordingly, some teachers may profit more from following a set of scripts than others, for example, unexperienced teachers. Scripted guidelines may also be a support in an introductory phase of an intervention for teachers who do not know their students when they start working with them, such as was the case for the sample of teachers in the present study. In this regard, this sample differed from previous studies in which students' classroom teachers usually were the main provider of the intervention (Rogde et al., 2016; Wasik et al., 2006; Yoshikawa et al., 2015). Working with an unfamiliar group of students adds additional challenges when implementing an intervention; challenges such as building relations with students over a short amount of time, and having limited information about students' background, knowledge, and level of proficiency. Thus, scripting can be the first step to improve the quality of instruction. Scripted intervention guidelines can be efficient as they require less preparation time and help teachers be consistent in implementing the program, and thus, may ease evaluations of implementation fidelity.

However, scripting may restrict teachers' professional autonomy and creativity, and reduce instructional quality if teachers become focused on following scripted guidelines rather than attending to individual students' needs (Bleses et al., 2018; Parks & Bridges-

Rhoads, 2012). Taking time to explain when misunderstandings occur and provide additional information are instructional qualities that may be neglected if teachers are too preoccupied with following and getting through a scripted program. Thus, strongly scripted guidelines can reduce the developmental appropriateness of teachers' interactions with students. Finally, the optimal balance between scripted and less scripted parts of an intervention may vary depending on educational cultures in specific countries. Scandinavian countries typically value child-oriented learning activities, and manualized programs are uncommon, emphasizing instead that teachers should decide on practices and methods that fit the group they work with. This may be one reason why the less scripted version of a Danish program was more effective than the scripted ones (Bleses et al., 2018).

An inclusion criterion in the study was that students demonstrated second-language skills in need of support (scored at least 1.5 SD below the mean). It is therefore reasonable to believe that the higher proportion of labeling talk categories in the early exchanges was helpful in providing a joint platform of understanding. Then, as teachers got to know their students better, and the students started talking more, the teachers apparently were more likely to demonstrate talk that was less scripted. To know a word, a learner needs to know the concept behind the label. Therefore, merely hearing the same word multiple times during a lesson is not enough to fully understand the concept, nor does it provide a richer language environment. According to the social-interactionist perspective on language learning that the present study is framed within, quality features of language develop through the dynamic interactions between teacher and students as questions and background experiences are brought into the conversation (Halliday, 1993; Ninio & Snow, 1996). Teachers' ability of being responsive to and following up on students' contributions promote language learning (Hansen, 2018). Recent research on quality of linguistic input has shown that the most impactful language practices were those targeting conceptual information, definitions, and

word relations (Bowne et al., 2016; Neugebauer et al., 2017; Silverman et al., 2014).

Moreover, the present study indicated that features building on conceptual definitions and

usage-based examples of these in combination with extended discourse can help teachers

create enriched language environments for second-language learners. Affording students with

semantic categories and semantically related words allows them to draw connections between

what they already know and relate it to words within the same topic. This aligns with studies

demonstrating the effectiveness of teaching children conceptually aligned categories (Hadley

et al., 2018; Neuman et al., 2011; Pinkham et al., 2014). Repeated exposure of this kind of

linguistic input, therefore, may offer an opportunity to go beyond the simple word to learn the

concept behind the words. These are all factors crucial in promoting second-language learners'

oral language skills. Using semantic categories can also make it easier for teachers to talk

about definitions and include examples that connect with children's' own experiences. This

may lead to topic-related stories and explanations offering children multiple opportunities to

engage in conversations.

Limitations

This study had limitations. First, the small sample size reduced the statistical power of

the analysis. Therefore, we were not able to explore whether variations across individual

teachers' instructional talk predicted child outcomes above the effect of receiving the program

that previously has been demonstrated ($d = .35$ when the intervention group was compared

with a control group; see Authors et al., 2019). Second, we were not able to reliably identify

individual children's utterances in the audiotaped observations of instructional talk, and

therefore, coding of the children's utterances was conducted at the group level. We cannot

exclude the possibility that the student groups' participation in labeling, definitions, and

extended discourse at both time points in some groups may have resulted from specific

students' contributions, while in other groups reflected a participation structure in which all

students joined. Video-taped observations would have allowed a more fine-grained analysis of teacher–student interactions at both time points, as well as the possible development of these interactions over the intervention period.

Conclusion

This study addressed characteristics of teachers' instructional talk in the early and late parts of an intensive second-language intervention that had demonstrated substantial effects on young second-language learners' oral language skills. We examined whether potential developments of teachers' instructional talk over the time span were aligned with intervention components that differed in degree of scriptedness. The findings suggested that developments in teachers' instructional talk, in particular, were demonstrated within talk categories aligned with the less scripted part of the intervention, while we found no changes in teachers' labeling talk that was associated with the more scripted parts of the intervention. Promoting oral language skills in students relies on teachers providing language-rich environments where word knowledge, vocabulary, and conceptual knowledge are incorporated in the teachers' instructional practices. Sparsity of concept-rich interactions in classrooms often results in fewer encounters with unfamiliar words underlying oral comprehension and reading comprehension, which are particularly important for second-language learners to reach proficiency. Very few studies have examined the extent to which teachers change their instructional talk during an intervention. This may have implications for the design of intervention programs. Future research should address how modifications in teachers' instructional talk in relation to scripting affect child outcomes. In particular, more emphasis should be placed on fine-grained methodology as the analysis of teacher–student interactions.

References

Aarts, R., Demir-Vegter, S., Kurvers, J., & Henrichs, L. (2016). Academic language in shared book reading: Parent and teacher input to mono- and bilingual preschoolers. *Language Learning, 66*(2), 263–295. doi:10.1111/lang.12152

Author. (2007).

Authors et al. (2019).

Author, & Colleagues. (2019).

Authors, & Colleagues. (2019).

Bleses, D., Højen, A., Dale, P. S., Piasta, S. B., Justice, L. M., Dybdal, L., . . . Haghish, E. F. (2018). Effective language and literacy instruction: Evaluating the importance of scripting and group size components. *Early Childhood Reading Quarterly, 42*(1), 256–269. doi:10.1016/j.ecresq.2017.10.002

Bowne, J. B., Yoshikawa, H., & Snow, C. E. (2016). Experimental impacts of a teacher professional development program in early childhood on explicit vocabulary instruction across the curriculum. *Early Childhood Research Quarterly, 34*, 27–39. doi:10.1016/j.ecresq.2015.08.002

Carlo, M. S., August, D., McLaughlin, B., Snow, C. E., Dressler, C., Lippman, D. N., …White, C. E. (2004). Closing the gap: Addressing the vocabulary needs of English-language learners in bilingual and mainstream classrooms. *Reading Research Quarterly*, *39*(2), 188–215. doi:10.158/RRQ.39.2.3

Colleagues, & Author. (2011).

Colleagues, & Author. (2014).

Dickinson, D. K. (2011). Teachers' language practices and academic outcomes of preschool children. *Science, 333*(6045), 964–967. doi:10.1126/science.1204526

Running head: TEACHERS' INSTRUCTIONAL TALK

Dickinson, D. K., & Caswell, L. (2007). Building support for language and early literacy in

 preschool classrooms through in-service professional development: Effects of the

 Literacy Environment Enrichment Program (LEEP). *Early Childhood Research*

 Quarterly, 2(2), 243–260. doi:10.1016/j.ecresq.2007.03.001

Dickinson, D. K., & Porche, M. V. (2011). Relation between language experiences in

 preschool classrooms and children's kindergarten and fourth-grade language and

 reading abilities. *Child Development*, *82*(3), 870–886. doi:10.1111/j.1467-

 8624.2011.01576.x

Elleman, A. M., Lindo, E. J., Morphy, P., & Compton, D. L. (2009). The impact of

 vocabulary instruction on passage-level comprehension of school-age children: A

 meta-analysis. *Journal of Research on Educational Effectiveness*, *2*(1), 1–44.

 doi:10.1080/19345740802539200

Hadley, E. B., Dickinson, D. K., Hirsh-Pasek, K., & Golinkoff, R. M. (2018). Building

 semantic networks: The impact of a vocabulary intervention on preschoolers' depth of

 word knowledge. *Reading Research Quarterly, 54*(1), 41–61. doi:10.1002/rrq.225

Halliday, M. A. K. (1993). Towards a language-based theory of learning. *Linguistics and*

 Education, 5(2)*,* 93–116. doi:10.1016/0898-5898(93)90026-7

Hansen, J.E. (2018). Characteristics of educational language practices observed in four

 toddler groups in Norwegian ECEC. *International Journal of Early Years Education*.

 doi:10.1080/09669760.2018.1548347

Harn, B., Parisi, D., & Stoolmiller, M. (2013). Balancing fidelity with flexibility and fit: What

 do we really know about fidelity of implementation in school? *Exceptional Children,*

 79(2), 181–193. doi:10.1177/001440291307900204

Hindman, A. H., & Wasik, B. A. (2012). Unpacking an effective language and literacy

coaching intervention in Head Start. *The Elementary School Journal, 113*(1), 131–

154. doi:10.1086/666389

Hoff, E. (2006). How social contexts support and shape language development.

Developmental Review, *26*(1), 55–88. doi:10.1016/j.dr.2005.11.002

Justice, J. M., Jiang, H., & Strasser, K. (2018). Linguistic environment of preschool

classrooms: What dimensions support children's language growth? *Early Childhood

Research Quarterly, 42*(1), 79–92. doi:10.1016/j.eresq.2017.09.003

Justice, L. M., Mashburn, A. J., Hamre, B. K., & Pianta, R. C. (2008). Quality of language

and literacy instruction in preschool classrooms serving at-risk pupils. *Early

Childhood Research Quarterly, 23*(1), 51–68. doi:10.1016/j.ecresq.2007.09.004

Kim, J. S., Burkhauser, M. A., Quinn, D. M., Guryan, J., Kingston, H. C., & Aleman, K.

(2017). Effectiveness of structured teacher adaptations to an evidence-based summer

literacy program. *Reading Research Quarterly, 52*(4), 443–467. doi:10.1002/rrq.178

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical

data. *Biometrics, 33*(1), 159–174.doi:10.2307/2529310

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah,

NJ: Erlbaum.

Marulis, L. M., & 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. doi:10.3102/0034654310377087

Michener, C. J., Proctor, C. P., & Silverman, R. D. (2018). Features of instructional talk

predictive of reading comprehension. *Reading and Writing, 31*(3), 725–756.

doi:10.1007/s11145-017-9807-4

Running head: TEACHERS' INSTRUCTIONAL TALK

Neugebauer, S., Coyne, M., McCoach, B., & Ware, S. (2017). Teaching beyond the

intervention: The contribution of teacher language extensions to vocabulary learning

in urban kindergarten classrooms. *Reading and Writing, 30*, 543–567.

doi:10.1007/s11145-016-9689-x

Neuman, S. B., & Cunningham, L. (2009). The impact of professional development and

coaching on early language and literacy instructional practices. *American Educational*

Research Journal, 46(2), 532–566. doi:10.3102/0002831208328088

Neuman, S. B., Newman, E. H., & Dwyer, J. (2011). Educational effects of vocabulary

intervention on preschoolers' word knowledge and conceptual development: A

cluster-randomized trial. *Reading Research Quarterly*, *46*(3), 249–272.

doi:10.1598/RRQ.46.3.3

Ninio, A., & Snow, C.E. (1996). *Pragmatic development*. Oxford, England: Westview Press.

Parks, N., & Bridges-Rhoads, S. (2012). Overly scripted: Exploring the impact of a scripted

literacy curriculum on a preschool teacher's instructional practices in mathematics.

Journal of Research in Childhood Education, 26(3), 308–324.

doi:10.1080/02568543.2012,684422

Pinkham, A. S., Kaefer, T., & Neuman, S. B. (2014). Taxonomies support preschoolers'

knowledge acquisition from storybooks. *Child Development Research, 2014,* 1–10.

doi:10.1155/2014/386762

Roberts, T., & Neal, H. (2004). Relationships among preschool English language learner's

oral proficiency in English, instructional experience and literacy development.

Contemporary Educational Psychology, *29*(3), 283–311.

doi:10.1016/j.cedpsych.2003.08.001

Rogde, K., Melby-Lervåg, M., & Lervåg, A. (2016). Improving the general language skills of

second-language learners in kindergarten: A randomized controlled trial. *Journal of*

Running head: TEACHERS' INSTRUCTIONAL TALK

> *Research on Educational Effectiveness, 9*(1), 150–170.
>
> doi:10.1080/19345747.2016.1171935

Silverman, R. D., Proctor, C. P., Harring, J. R., Doyle, B., Mitchell, M. A., & Meyer, A. G.

> (2014). Teachers' instruction and students' vocabulary and comprehension: An
>
> exploratory study with English monolingual and Spanish-English bilingual students in
>
> grades 3-5. *Reading Research Quarterly, 49*(1), 31–60. doi:10.1002/rrq.63

Smolkowski, K., & Gunn, B. (2012). Reliability and validity of the classroom observations of

> student-teacher interactions (COSTI) for kindergarten reading instruction. *Early*
>
> *Childhood Research Quarterly, 27*, 316–328. doi:10.1016/j.ecresq.2011.09.004

Snow, C. E., Tabors, P. O., & Dickinson, D. K. (2001). Language development in the

> preschool years. In D. K. Dickinson & P. O. Tabors (Eds.), *Beginning literacy with*
>
> *language: Young children learning at home and school* (pp. 1–25). Baltimore, MD:
>
> Brookes.

Wasik, B. A., Bond, M. A., & Hindman, A. (2006). The effects of a language and literacy

> intervention on Head Start children and teachers. *Journal of Educational Psychology,*
>
> *98*(1), 63–74. doi:10.1037/0022-0663.98.1.63

Wasik, B. A., & Hindman, A. H. (2011). Improving vocabulary and pre-literacy skills of at-

> risk preschoolers through teacher professional development. *Journal of Educational*
>
> *Psychology, 103*(2), 455–469. doi:10.1037/a0023067

Wasik, B. A., & Hindman, A. H. (2018). Increasing preschoolers' vocabulary development

> through a streamlined teacher professional development intervention. *Early Childhood*
>
> *Research Quarterly*. doi:10.1016/j.ecresq.2018.11.001

Yoshikawa, H., Leyva, D., Snow, C. E., Trevino, E., Barata, M. C., Weiland, C., . . . Arbour,

> M. C. (2015). Experimental impacts of a teacher professional development program in

Chile on preschool classroom quality and child outcomes. *Developmental Psychology,*

51(3), 309–322. doi:10.1037/a0038785

Zucker, T. A., Justice, L. M., Piasta, S. B.., & Kaderavek, J. N. (2010). Preschool teachers'

literal and inferential questions and children's responses during whole-class shared

reading. *Early Childhood Research Quarterly, 1*(1), 65-83.

doi:10.1016/j.ecresq.2009.07.001

Table 1

Teachers' Talk Categories per Minute

Variable name		Time 1						Time 2			
	N	M	SD	Min	Max	M	SD	Min	Max		
Simple labeling	15	4.72	2.48	1.99	11.93	3.60	1.98	1.63	7.92		
Elaborated labeling	15	1.10	.65	.23	2.06	.75	.47	.10	1.71		
Conceptual definitions	15	.16	.16	.00	.56	.40	.33	.00	1.05		
Usage-based examples of definitions	15	.17	.19	.00	.62	.58	.49	.00	1.89		
Extended discourse	15	.58	.63	.00	1.75	3.10	1.99	.46	7.41		

Table 2

Student Groups' Talk Categories per Minute

Variable name	N	Time 1				Time 2			
		M	SD	Min	Max	M	SD	Min	Max
Simple labeling	15	6.90	3.00	2.59	13.95	5.39	2.32	1.90	10.18
Elaborated labeling	15	.54	.42	.00	1.33	1.14	.75	.07	2.96
Conceptual definitions	15	.02	.05	.00	.20	.07	.10	.00	.32
Usage-based examples of definitions	15	.16	.19	.00	.76	.66	.63	.00	2.12
Extended discourse	15	.46	.55	.00	1.84	2.46	1.63	.10	5.07

Kartlegging av språkferdigheter for elever på 1.-4.trinn: En vurdering av kartleggingsverktøyet Norsk Som Læringsspråk (NSL)

Mia C. Heller og Arne Lervåg

Universitetet i Oslo

Sammendrag

Kartlegging av språkferdigheter kan bidra til tidlig identifisering av barn som trenger støtte i skolen og dermed være veiledende for innholdet i undervisningen. Kvaliteten ved måleinstrumentene er viktig ved kartlegging. Denne studien gjennomførte analyser av reliabilitet, begrepsvaliditet og kriterievaliditet i kartleggingsverktøyet «Norsk Som Læringsspråk» (NSL), som er en læreplanorientert språkkartlegging for 1.-4.trinn. NSL vurderer ferdigheter innenfor kategorisering, vokabular (bredde og dybde) og grammatikk. Konfirmerende faktoranalyser identifiserte NSL som en fire-faktormodell. Det ble også funnet korrelasjoner mellom NSL og andre standardiserte språkkartlegginger som underbygger validiteten i kartleggingsverktøyet.

Summary

Language assessment can be a valuable source of information to identify students in need of language support. An assessment tool's usability for educational purposes largely depends on using results to adapt instruction that can benefit student's learning. For an assessment tool to become a valuable source of information it must be both reliable and valid. This study conducted an examination of reliability, construct validity and criterion validity of the assessment tool "Norwegian as a Language for Learning" (NSL). The NSL is a standardized curriculum-based language test for students in grades 1-4 and assess skills related to categorization, vocabulary (breadth and depth), and grammar. Confirmatory factor analysis identified NSL as a four-factor model. Furthermore, correlations between NSL and other standardized language tests were found, which supports its validity.

Innledning

Gode språkferdigheter er et viktig grunnlag for læring og deltakelse i et sosial

fellesskap. I dagens flerkulturelle skole er det en økende andel elever som har et annet

morsmål enn norsk. Begrepet minoritetsspråklig defineres ofte i skolesammenheng som barn

som har et annet førstespråk enn norsk eller samisk (Opplæringslovens § 2-8). Tall for 2018

(Statistisk sentralbyrå, 2019) viser at andelen barn med vedtak om særskilt norskopplæring

utgjør 6,7 % av elevtallet i grunnskolen i Norge. I Oslo utgjør denne gruppen elever 19,5 %.

Studier viser at minoritetsspråklige elever som gruppe har svakere ferdigheter på skolens

språk enn sine enspråklige medelever (Melby-Lervåg & Lervåg, 2014). Dette stiller ulike krav

både til skolen og den enkelte elev i forhold til språkkompetanse, og det tydeliggjør

utfordringene når det gjelder tilrettelegging i skolen. Kunnskap om hva språk er og hvordan

språkferdigheter utvikles og måles er avgjørende for å identifisere minoritetsspråklige elever

som trenger språkstøtte. Betydningen av barns norskspråklige ferdigheter og hvordan

barnehagen og skolen kan bidra til å støtte barnas språkutvikling løftes frem i

stortingsmeldingen om tidlig innsats og inkluderende fellesskap (Kunnskapsdepartementet,

2019). Innenfor en norsk kontekst finnes det foreløpig få kartleggingsverktøy som skolen kan

benytte for vurdering av språkferdigheter. Det er ingen føringer for hvilke

kartleggingsverktøy som skal ligge til grunn utover de nasjonale og kommunale prøvene som

er obligatoriske. Disse kartleggingene måler barnas leseferdigheter. Det kan derfor være en

utfordring for skolen å sette inn relevante tiltak for å styrke språkutviklingen hos barn som

viser mangelfulle språkferdigheter.

Mangelen på verktøy som måler språkferdigheter var en av grunnene til utviklingen av

Norsk Som Læringsspråk (NSL; Frøyen, Ahmadinia, Heller, & Skjåk, 2011; Frøyen,

Ahmadinia, Heller, Skjåk, & Namvar, 2015). Man så behov for et instrument som kartla barns

språkferdigheter uten å stille krav til lese- og skriveferdigheter. Denne artikkelen vurderer

komponentene NSL består av gjennom analyser av reliabilitet, begrepsvaliditet og kriterievaliditet, med påfølgende implikasjoner for praksis.

Språkets betydning for skolefaglig utvikling

Språk er et komplekst fenomen. Det deles ofte i muntlige ferdigheter (ekspressive) og lytteferdigheter (impressive), i tillegg til komponentene semantikk (kunnskap om ord og setninger), grammatikk (syntaks og morfologi), fonologi (språklyder) og pragmatikk (bruk av språk). Disse komponentene påvirker hverandre gjensidig og gjennom innflytelser fra omgivelsene. Barn utvikler språket gjennom samspill, både som aktive deltakere og lyttere. Vokabular danner grunnlaget for språkutviklingen gjennom muligheter for å kommunisere tanker, meninger, følelser og behov, men også ved evnen til å systematisere og kategorisere ord i systemer. Det skilles ofte mellom vokabularbredde og vokabulardybde. Mens førstnevnte viser til antall ord barn har kjennskap til (for eksempel gjennom peking og benevning av bilder), viser vokabulardybden til kunnskapen barnet har om ord (for eksempel å definere et ord). Ordinnlæringen henger nært sammen med barns grammatiske utvikling ved at evnen til å sette sammen ord til enkle ytringer forutsetter at en viss mengde ord er lært. Antall ord et barn forstår og kan bruke er til hjelp for å hente ut mening av en ytring. Samtidig kan en ytring bidra med kontekstuell støtte for å forstå nye ord eller bidra til dypere kunnskap om et ord. Denne parallelle utviklingen og relasjonen mellom vokabular og grammatikk bekreftes gjennom signifikant høye korrelasjoner i studier av språkutvikling (Klem et al., 2015; Lervåg, Hulme, & Melby-Lervåg, 2018).

Tvillingstudier viser at gener påvirker språktilegnelsen (Hayiou-Thomas, Dale, & Plomin, 2012), samtidig som miljøet barnet er en del av spiller en sentral rolle (Hart & Risley, 1995). Mengden og variasjonen av språkstimuleringen barn blir eksponert for i tidlig alder har sammenheng med senere språkferdigheter (Hoff, 2006; Rowe, 2012). Samtidig forekommer

det forskjeller i språkstimuleringen med bakgrunn i de sosioøkonomiske forholdene barn vokser opp i. Det etableres tidlig forskjeller i størrelsen på barns ordforråd (Hart & Risley, 1995). Disse forskjellene har vist seg å vedvare over tid (Melby-Lervåg et al., 2012; Storch & Whitehurst, 2002). Dette er bekymringsfullt sett i lys av at skolerelaterte ferdigheter i stor grad predikeres av barnets ordforråd i førskolealder (Dickinson & Tabors, 2001; Dickinson & Porche, 2011; Hjetland, Brinchmann, Scherer, & Melby-Lervåg, 2017).

Barn møter allerede fra skolestart et språk preget av abstraksjon og som kjennetegnes av et avansert vokabular, kompleks syntaks og krav til varierte pragmatiske ferdigheter (Snow & Uccelli, 2009). Dette er ferdigheter som ikke er temaspesifikke, men som gir muligheter for å forstå faginnhold på tvers av fagområder. Ferdighetene er betydningsfulle for tilegnelse av kunnskap, spesielt med tanke på at det akademiske språket også er undervisningsspråket i skolen. I lys av dette stilles det uforholdsmessige store krav til hva barn som er på vei til å tilegne seg et andrespråk skal få med seg i løpet av skoledagen. Dette fordi elevene både er i en prosess hvor et nytt språk skal læres og mestres, samtidig som den skolefaglige kunnskapen blir formidlet og skal tilegnes gjennom et språk de foreløpig ikke behersker godt nok.

Hvordan kan språk måles?

Måling av språkferdigheter stiller krav til en operasjonalisering av begrepet språk. Dette innebærer å gjøre begrepet målbart ved bruk av ulike kartleggingsverktøy som representerer de språklige komponentene. Kartlegging av barn som trenger språkstøtte handler ofte om å identifisere hvilke ord (vokabular) og setninger (grammatikk) et barn forstår og kan bruke, samt forteller ferdigheter (pragmatikk). Det er vanlig å bruke standardiserte kartleggingsverktøy for å sammenlikne barns språkferdigheter med hva som er forventet for alderen. Eksempler på slike kartleggingsverktøy er British Picture Vocabulary Scale (BPVS-

II; Dunn, Dunn, Whetton, & Burley, 1997; Lyster, Horn, & Rygvold, 2010) som vurderer impressivt vokabular og Test for Reception of Grammar (TROG-2; Bishop, 2003; Lyster & Horn, 2009) som måler grammatisk forståelse. Selv om hensikten med kartleggingsverktøy er å gi presise målinger av begrepet språk, vil det forekomme målefeil. Målefeil kan være systematiske (skjev måling av begrepet på grunn av for snever eller for bred operasjonalisering) eller tilfeldige (tilfeldigheter man ikke kan kontrollere slik som elevens dagsform under kartleggingen). Det er derfor viktig med informasjon om komponentene kartleggingsverktøy består av.

Et springende punkt ved utvikling av kartleggingsverktøy handler om instrumentets kvalitet. Dette vurderes ut ifra spesifikke kriterier for validitet og reliabilitet, som baseres på analyser av komponentene instrumentet består av. Reliabilitet gir informasjon om hvor stabilt og reproduserbart måleinstrumentet er (målesikkerhet) (Thorndike & Thorndike-Christ, 2014). Validitet uttrykker i hvilken grad teorien og empirien som ligger til grunn støtter tolkningen av testresultatene innenfor en spesifikk kontekst (Messick, 1995). For at et kartleggingsverktøy skal ha en nytteverdi, både for forskning og praksisfeltet, må det være reliabelt og valid. Dersom en test viser seg å ha lav reliabilitet, vil den heller ikke egne seg til bruk da den ikke er valid (gyldig). Samtidig er det viktig å påpeke at god reliabilitet ikke er en garanti for validitet, da en test kan vise seg å være pålitelig uten at den faktisk måler det den er utviklet for å måle (Thorndike & Thorndike-Christ, 2014). I denne artikkelen fokuseres det på begrepsvaliditet og kriterievaliditet i evalueringen av validitet i kartleggingsverktøyet NSL. Begrepsvaliditet dreier seg om i hvilken grad komponentene måleinstrumentet består av er representative for begrepet de er ment å måle. Ved utvikling av et kartleggingsverktøy vurderes begrepsvaliditet gjennom estimering av faktormodeller (Brown, 2015). Faktoranalyser undersøker hva som er felles variasjon mellom et sett av variabler for å vurdere strukturen i kartleggingsverktøyet. Analysene gir informasjon om begrepet, i dette

tilfellet språk, er best forklart som en felles dimensjon eller om det representerer separate dimensjoner. Flere studier finner støtte for at variasjon i barns språkferdigheter, da spesielt vokabular, grammatikk og språkbruk, kan forklares gjennom en endimensjonal språkfaktor (Klem et al., 2015; Tomblin & Zhang, 2006). Siden normer brukes som retningslinjer for tolkning av elevers prestasjoner, er det viktig å vite hvorvidt kartleggingsverktøyet er representativt for barnegruppen det brukes på. Dette handler om invarians, i hvilken grad det operasjonaliserte begrepet kan sies å måle det samme på tvers av grupper eller over tid (Kline, 2016). I forskningslitteraturen vises det til grad av invarians, og dette måles basert på faktorstrukturen (nivå 1; konfigural), faktorladningene (nivå 2; metrisk) og intersept gjennomsnittet (nivå 3; skalar). Det laveste nivået, faktorstrukturen, må være oppfylt for å kunne vurdere invarians over faktorladningene som igjen må være oppfylt for kunne vurdere invarians over interseptene (Putnik & Bornstein, 2016). Skalar invarians er en forutsetning for å kunne sammenlikne gjennomsnitt mellom grupper (Kline, 2016). Det brukes ofte for å sammenlikne om det operasjonaliserte begrepet (språk) måles på samme måte i forhold til alder, kjønn, sosioøkonomisk bakgrunn og språkstatus. Dersom det ikke er skalar invarians må det brukes forskjellige normer, da begrepet ikke måles likt mellom gruppene.

Kriterievaliditet vurderer hvorvidt måleinstrumentet samsvarer med andre instrumenter som måler det samme begrepet, og om instrumentet tilfører ny informasjon som ikke fanges opp av andre verktøy som allerede benyttes for samme formål. I tillegg belyses hvor godt måleinstrumentet klarer å identifisere elevgruppen det ble designet for (Thorndike & Thorndike-Christ, 2014). Kriterievaliditet kan enten være samtidig eller prediktiv. Dette avhenger av hvorvidt testene er foretatt på samme tidspunkt eller benyttes for å forutsi fremtidige prestasjoner.

Innenfor en norsk kontekst finnes det få studier av kvaliteten ved måleinstrumenter som benyttes i skolen for å vurdere barns skolefaglige ferdigheter. Arnesen, Braeken, Ogden

og Melby-Lervåg (2019) gjennomførte imidlertid nylig en undersøkelse som belyste

kvaliteten på måleinstrumenter brukt til kartlegging av sosial fungering og leseferdigheter i

barneskolen, og hvorvidt resultatene ble benyttet som føringer for innholdet i undervisningen.

Resultatene tilknyttet kvaliteten ved måleinstrumentene viste at mange hadde svak validitet

eller at det ikke var foretatt noen kvalitetsvurdering av måleinstrumentet. Dette er spesielt

bekymringsfullt sett i lys av at flere av disse kartleggingsverktøyene er hyppig brukt i skolen.

Når det gjaldt nytteverdien av måleinstrumentet, i hvilken grad resultatene var rettledende for

undervisningen, indikerte funn at føringer for praksis i stor grad var basert på lærernes

uformelle vurderinger fremfor informasjonen som fremkom av resultatene fra kartleggingene.

Formålet med studien

Intensjonen med denne studien var å gjennomføre en vurdering av komponentene i

NSL (Frøyen et al., 2011; 2015) og grad av samsvar med øvrige språkkartlegginger.

Komponentene i NSL baseres på teoretiske og praktiske tilnærminger til språk for tilegnelse

av kunnskap. Den teoretiske tilnærmingen baseres på teorier og studier om språkutvikling,

både generell og tospråklig språkutvikling. Dette lå til grunn for utvelgelse av komponentene i

NSL (kategorisering, vokabular og grammatikk). Den praktiske tilnærmingen tok

utgangspunkt i at kartleggingsverktøyet skulle være lærerplanorientert ved at det var foretatt

analyser av hvilke krav som stilles til elevenes språkferdigheter i Kunnskapsløftet (LK06;

Utdanningsdirektoratet, 2006). Resultatene ble kombinert med analyser av ord fra lærerbøker

i fagene norsk, matematikk, samfunnsfag, naturfag og KRLE. Disse analysene lå til grunn for

utvelgelsen av ordene i NSL. Dette var ord barn nødvendigvis ikke eksponeres for gjennom

dagligdagse samtaler.

Hensikten med NSL er å identifisere elever i behov av å styrke norskspråklige

ferdigheter i barneskolen (1.-4.trinn), for deretter å tilrettelegge opplæringen basert på

resultatene som foreligger. Den første versjonen av NSL ble ferdigstilt i 2011.

Kartleggingsverktøyet hadde da vært igjennom en pilotering (94 barn ble kartlagt), hvor

sammenhenger mellom resultater på NSL opp mot resultater på de standardiserte språktestene

Test for Reception of Grammar, 2nd edition (TROG-2; Bishop, 2003; Lyster & Horn, 2009)

og delprøven Ordforståelse fra Wechsler Intelligence Scale for Children, 4th edition (WISC-

IV; Wechsler, 2003) ble undersøkt for å måle kriterievaliditet. Det ble funnet sterke

korrelasjoner, både mellom NSL og TROG-2 (r = .70) og mellom NSL og Ordforståelse (r

= .79). NSL ble deretter normert på 245 barn. Etter at NSL hadde vært i bruk i over tre år, ble

det foretatt en ny normering. Denne renormeringen var hovedsakelig basert på

tilbakemeldinger fra praksisfeltet og med et formål om å supplere antall elever i

normeringsgrunnlaget. Den nye versjonen av NSL ble ferdigstilt i 2015 (Frøyen et al., 2015).

I dette arbeidet ble det ikke foretatt nye sammenlikninger med andre standardiserte

språktester. NSL ble imidlertid brukt for å identifisere minoritetsspråklige barn i behov av

språkstøtte til en intervensjonsstudie, hvor det også ble benyttet andre

språkkartleggingsverktøy for å vurdere barnas ferdigheter (Heller, Lervåg, & Grøver, 2019).

Ved å inkludere utvalget fra intervensjonsstudien sammen med normeringsutvalget til NSL

fra 2014, fokuserer denne studien spesielt på å besvare følgende forskningsspørsmål:

1. I hvilken grad er den indre konsistensen i NSL tilfredsstillende (reliabilitet)?

2. I hvilken grad er strukturen i NSL best forklart gjennom en en-faktormodell eller en
   flerfaktormodell (begrepsvaliditet)?

3. I hvilken grad viser NSL invarians over språkstatus, kjønn og klassetrinn?

4. I hvilken grad er det samsvar mellom NSL og andre standardiserte språkkartlegginger
   (kriterievaliditet)?

Metode

Deltakere

I denne studien ble det benyttet to utvalg. Det første utvalget var normeringsgrunnlaget til NSL fra 2014 som bestod av 373 elever (1.-4.trinn) fordelt på 41 skoler. 60.1 % av barnegruppen var enspråklige (47.5 % jenter, gjennomsnittsalder = 92.52 måneder, *SA* = 14.00 måneder). Denne selekteringen var basert på at utvalget skulle gjenspeile den flerkulturelle populasjonen i skolen. Det ble gjennomført et tilfeldig utvalg av åtte skoler fra hver skolegruppe i Oslo, med krav om at andelen minoritetsspråklige elever utgjorde 20-40 % av utvalget. Åtte elever fra hver skole fordelt på to trinn (1. og 3.trinn eller 2. og 4.trinn) ble tilfeldig valgt ut etter innhenting av foresattes samtykke. Totalt 40 minoritetsspråk var representert hvor urdu, engelsk, arabisk, somali, vietnamesisk og spansk utgjorde 44.1 %. Åtte barn ble kartlagt på hver av skolene, enten kombinasjonen 1. og 3.trinnselever eller 2. og 4.trinnselever. Ingen av barna hadde vedtak om spesialundervisning (Opplæringslovens § 5-1), og ingen av de minoritetsspråklige barna hadde vedtak om særskilt norskopplæring (Opplæringslovens § 2-8).

Utvalg II bestod av 137 barn (1. og 2.trinn) fordelt på 16 skoler, alle minoritetsspråklige elever (54 % jenter, gjennomsnittsalder = 75.34 måneder, *SA* = 6.20 måneder). I dette utvalget ble NSL benyttet som screener for å identifisere elever med mangelfulle norskspråklige ferdigheter for deltakelse i en randomisert kontrollert studie som vurderte effekten av en språkintervensjon (Heller et al., 2019). Minoritetsspråklige barn som skåret 1.5 standardavvik eller mer under gjennomsnittet på NSL ble invitert til å delta i studien. Åtte til tolv barn fra hver av skolene ble tilfeldig valgt ut etter innhenting av foresattes samtykke. Totalt 31 språk var representert, hvor urdu, arabisk, somali, kurdisk, tyrkisk og tamilsk utgjorde 67.9 %. Alle elevene kvalifiserte for særskilt norskopplæring etter Opplæringslovens § 2-8.

Måleinstrumenter

Norsk Som Læringsspråk (NSL; Frøyen et al., 2011, 2015) består av 195 testledd

fordelt på 10 delprøver og måler elevens ferdigheter innenfor kategorisering, vokabular

(bredde og dybde) og grammatikk. Dette er oppgaver som stiller krav til språkforståelse

(impressive ferdigheter) og muntlige språkferdigheter (ekspressive ferdigheter).

Kategorisering vurderer evnen til å se relasjoner og systematisere ord ved at barnet peker på

hvilke bilder som hører sammen. Dette måles gjennom delprøvene matching av kategorier

(MK, 45 testledd), generalisering av kategorier (GK, 20 testledd). Vokabular måler størrelsen

på vokabularet gjennom peking (impressiv) og benevning av bilder (ekspressiv), med andre

ord bredden i elevens ordforråd. Dette vurderes gjennom delprøvene ordforståelse impressiv

(OFI, 29 testledd), ordforståelse ekspressiv (OFE, 39 testledd). Antonymer/synonymer stiller

krav til å sammenlikne likheter og forskjeller mellom ord på et mer abstrakt nivå, og måler

dybden i ordforrådet ved at barnet skal finne det motsatte ordet eller et ord som betyr det

samme som det som blir opplest. Dybdevokabularet måles gjennom delprøvene antonymer

(ANT, 14 testledd), synonymer (SYN, 16 testledd). Grammatikk vurderer setningsforståelse

(impressiv), hvor eleven får opplest en og en setning for så å peke ut riktig bilde til setningen.

Dette måles gjennom delprøvene adjektiv (ADJ, 13 testledd), preposisjoner (PRE, 7 testledd),

pronomen (PRO, 7 testledd) og adverb (ADV, 5 testledd). Alle delprøvene i NSL er

bildebasert med unntak av synonymer. Et poeng gis for hvert riktig svar og null poeng for feil

svar. Kartleggingen tar i gjennomsnitt 30 minutter og administreres individuelt.

British Picture Vocabulary Scale, 2nd edition (BPVS-II; Dunn, Dunn, Whetton, &

Burley, 1997; Lyster, Horn, & Rygvold, 2010). Dette er en impressiv språktest hvor barnet får

i oppgave å identifisere et ord som blir lest opp av testleder ved å peke ut riktig bilde basert på

fire alternativer. Den norske utgaven av BPVS-II består av 144 testledd.

Clinical Evaluation of Language Fundamentals, 4th edition (CELF-4; Semel, Wiig, & Secord, 2003). Delprøven *Ekspressivt Ordforråd* kartlegger elevens muntlige benevningsferdigheter. Testleder viser et og et bilde som barnet deretter skal benevne. Delprøven har 20 testledd.

Wechsler Preschool and Primary Scale of Intelligence, 4th edition (WPPSI-IV; Wechsler, 2012)/*Wechsler Intelligence Scale for Children, 4th edition* (WISC-IV; Wechsler, 2003). Delprøven *Ordforståelse* måler elevens evne til å gi en definisjon eller en forklaring av et ord (ekspressiv). Testleder leser opp et og et ord for barnet. Denne delprøven inneholder 30 testledd. Med bakgrunn i barnas språkferdigheter, ble en kombinasjon bestående av ord fra *Ordforståelse* i WPPSI-IV og WISC-IV benyttet. Denne versjonen har tidligere blitt brukt i norske studier som vurderte effekten av språkintervensjoner (Hagen, Melby-Lervåg, & Lervåg, 2017; Rogde, Melby-Lervåg, & Lervåg, 2016).

Test for Reception of Grammar, 2nd edition (TROG-2; Bishop, 2003; Lyster & Horn, 2009) er en impressiv språktest som kartlegger elevens forståelse av grammatikk. Setninger blir lest opp en etter en, og barnets oppgave er å peke ut riktig bilde basert på fire alternativer. Testen består av 80 testledd.

Bus Story (Renfrew, 1991) er en narrativ test som måler elevens evne til å gjenfortelle en historie etter å ha fått den opplest (ekspressiv). Selve historien blir fortalt av testleder ved bruk av bildestøtte (12 bilder). Testen innehar 32 testledd.

Prosedyre

Datagrunnlaget for denne studien ble innhentet over to perioder, høsten 2014 og høsten 2016. Elevene i utvalg I ble kartlagt med NSL høsten 2014. Kartleggingen ble foretatt på skolen og gjennomført av pedagogisk-psykologiske rådgivere med testkompetanse. Samtlige testledere hadde gjennomgått opplæringskurset i NSL og benyttet

kartleggingsverktøyet i praksis i forkant av datainnsamlingen. Utvalg II ble kartlagt med NSL

og de fem standardiserte språktestene (BPVS-II, Ekspressivt Ordforråd fra CELF-4,

Ordforståelse fra WPPSI-IV/WISC-IV, TROG-2 og Bus Story) høsten 2016. Kartleggingen

gjennomført på skolen av førsteforfatter, som tidligere har jobbet som pedagogisk-

psykologisk rådgiver.

Analyser

Råskårer ble benyttet i alle analysene. For å besvare det første forskningsspørsmålet,

ble det først foretatt analyser av den indre konsistensen (samvariasjon mellom itemene) i NSL

ved bruk av Cronbachs alpha (verdier over .70 vurderes som tilfredsstillende; Kline, 2016).

Det andre forskningsspørsmålet ble adressert gjennom konfirmerende faktoranalyser for å

undersøke strukturen i NSL. Analysene ble utført i M*plus* versjon 8 (Muthén & Muthén,

2017). Resultatene ble deretter analysert basert på kriterier for faktorladninger

(faktorladninger på størrelse med eller større enn .30 eller .40 blir ofte regnet som

betydningsfulle) og fit-indekser som gir informasjon om i hvilken grad den estimerte

faktormodellen passer til de observerte dataene (Brown, 2015). Hu og Bentler (1999)

anbefaler verdier rundt .06 som tilfredsstillende på the Root Mean Square of Error of

Approximation (RMSEA), Comparative Fit Index (CFI) og Tucker Lewis Index (TIL) bør

ligge på .95 eller høyere, mens the Standardized Root Mean Square Residual (SRMR) bør

være mindre enn .08. Kji-kvadrat differanse test ble brukt for å teste forskjellen mellom

faktormodellene. Den tester hvorvidt korrelasjonene mellom faktorene er signifikant

forskjellig fra 1. Siden kji-kvadrat differanse test er sensitiv for utvalgsstørrelse, ble

vurderingen av faktormodellene supplert med Bayesian informasjonskriterium (BIC). BIC

brukes for å sammenlikne modeller, hvor modellen med lavest BIC verdi har best tilpasning

(Kline, 2016). Det tredje forskningsspørsmålet ble besvart gjennom analyser av invarians ved

bruk av flergruppe konfirmerende faktoranalyser. For 2014-utvalget ble det gjennomført

analyser for språkstatus, kjønn og alderstrinn, mens det for 2016-utvalget kun var mulig å

analysere invarians for kjønn på grunn av utvalgsstørrelsen (78.8 % var 1.trinnselever). Til

slutt ble det foretatt en sammenlikning av NSL og de fem standardiserte språktestene (BPVS-

II, Ekspressivt Ordforråd fra CELF-4, Ordforståelse fra WPPSI-IV/WISC-IV, TROG-2 og

Bus Story) gjennom korrelasjonelle analyser for å besvare det fjerde forskningsspørsmålet om

kriterievaliditet.

Resultater

Deskriptiv informasjon om gjennomsnitt, standardavvik og reliabilitet er fremstilt i

tabell 1, mens tabell 2 viser korrelasjonene mellom delprøvene i NSL. Som det fremkommer

av tabell 2 er det forskjeller mellom de to utvalgene når det gjelder korrelasjoner mellom

delprøvene i NSL. Delprøvene matching av kategorier og generalisering av kategorier

korrelerer i liten grad med de øvrige delprøvene i 2016-utvalget sammenliknet med 2014-

utvalget. Når det gjelder det første forskningsspørsmålet, i hvilken grad den indre

konsistensen i NSL er tilfredsstillende, viser resultatene tilfredsstillende indre konsistens for

begge utvalgene med Cronbachs alfa over .70 (se tabell 1).

[Tabell 1 inn her]

[Tabell 2 inn her]

I hvilken grad er strukturen i NSL best forklart gjennom en en-faktormodell eller en fler-

faktormodell (begrepsvaliditet)?

For å kunne vurdere strukturen i NSL, ble det først estimert faktormodeller for 2014-

utvalget, hvor en en-faktormodell for språk bestående av alle de 10 delprøvene i NSL ble

estimert. Modelltilpasningen for en-faktormodellen var god (χ^2(29, N = 373) = 65.868,

p<.0012, CFI = .979, TLI = .972, SRMS = .030, og RMSEA = .049 (90% CI = .030-.067)).

Dette betyr at indikatorene, de ti delprøvene i NSL, er meningsfullt relatert til faktoren. Denne

en-faktormodellen ble så sammenliknet med en to-faktormodell (kategorisering og språk),

deretter en tre-faktormodell (kategorisering, vokabular og grammatikk) og til slutt en fire-

faktormodell (kategorisering, vokabular, antonymer/synonymer og grammatikk) (se tabell 3).

Modellsammenlikningen basert på fit-indeksene og Bayesian informasjonskriterium indikerte

at fire-faktormodellen bestående av faktorene kategorisering, vokabular,

antonymer/synonymer og grammatikk hadde best modelltilpasning til dataene (χ^2(29, N =

373) = 31.520, p<.3413, CFI = .998, TLI = .997, SRMS = .023, og RMSEA = .015 (90% CI

= .000-.043)). Kji-kvadrat differanse testing bekreftet dette ved at fire-faktormodellen var

signifikant forskjellig fra de tre andre faktormodellene (p = <.05). Figur 1 viser at

faktorladningene ligger godt over kriteriene (på størrelse med eller større enn .30 eller .40)

med verdier fra .55 til .91. Korrelasjonene mellom faktorene i fire-faktormodellen varierte

fra .82 til .98.

De samme faktormodellene ble deretter estimert for 2016-utvalget (se tabell 4). I

motsetning til 2014-utvalget, viste en-faktormodellen svak modelltilpasning (χ^2(29, N = 137)

= 82.871, p<.000, CFI = .880, TLI = .845, SRMS = .066, og RMSEA = .100 (90 % CI

= .072-.128)). Sammenlikningen av fit-indeksene og Bayesian informasjonskriterium til en-

faktormodellen med de tre øvrige faktormodellene, indikerte at også her hadde fire-

faktormodellen den beste modelltilpasningen (χ^2(29, N = 137) = 48.123, p<.0140, CFI = .952,

TLI = .925, SRMS = .044, og RMSEA = .070 (90 % CI = .031-.103)). Kji-kvadrat differanse

testing bekreftet dette ved at fire-faktormodellen var signifikant forskjellig fra de tre andre

faktormodellene (p = <.05). Som det fremkommer av figur 2, ligger faktorladningene over

kriteriene (på størrelse med eller større enn .30 eller .40) med verdier fra .38 til .82.

Korrelasjonene mellom faktorene varierte fra .10 til .79, hvor faktoren kategorisering ikke

korrelerte med vokabular.

[Tabell 3 inn her]

[Tabell 4 inn her]

[Figur 1 inn her]

[Figur 2 inn her]

I hvilken grad viser NSL invarians over språkstatus, kjønn og klassetrinn (begrepsvaliditet)?

Invarians, i hvilken grad begrepet måles på samme måte over tid eller grupper, ble

estimert for hvert av utvalgene. For 2014-utvalget ble det funnet skalar invarians for både

kjønn ($\Delta\chi^2(6) = 8.98$, $p = .174$) og språkstatus ($\Delta\chi^2(6) = 11.369$, $p = .078$), men ikke

klassetrinn ($\Delta\chi^2(12) = 55.571$, $p = .000$). For 2016-utvalget ble det funnet metrisk invarians

for kjønn ($\Delta\chi^2(6) = 7.202$, $p = .303$), men ikke skalar invarians ($\Delta\chi^2(6) = 17.789$, $p = .007$).

Funn av skalar invarians i 2014-utvalget gjorde det derfor mulig å sammenlikne

gjennomsnittene for kjønn og språkstatus. Barnas språkstatus (enspråklig eller

minoritetsspråklige) viste statistisk signifikante forskjeller i favør av de enspråklige elevene

på vokabular ($d = -.895$, $p = .000$), antonymer/synonymer ($d = -.446$, $p = .000$) og grammatikk

($d = -.448$, $p = .000$), mens det ikke forekom signifikante forskjeller knyttet til kategorisering

($d = -.203$, $p = .092$). Enspråklige barn gjorde det bedre på ordforråd og setningsforståelse,

mens det ikke var signifikante forskjeller mellom enspråklige og minoritetsspråklige elever

når det gjaldt kategorisering, en ferdighet som både stiller krav til logisk resonnering og

språk. Det var ingen signifikante forskjeller mellom jenter og gutter verken på kategorisering

($d = -.009$, $p = .949$), vokabular ($d = -.045$, $p = .722$), antonymer/synonymer ($d = -.038$, $p = .740$) eller grammatikk ($d = -.151$, $p = .196$).

I hvilken grad er det samsvar mellom NSL og standardiserte språkkartlegginger (kriterievaliditet)?

Tabell 5 viser korrelasjonene mellom firefaktormodellen til 2016-utvalget og de fem standardiserte språktestene (BPVS-II, Ekspressivt Ordforråd fra CELF-4, Ordforståelse fra WPPSI-IV/WISC-IV, TROG-2 og Bus Story). Som det fremkommer av tabellen, er det moderate til sterke korrelasjoner mellom faktoren vokabular og de standardiserte språktestene, samt mellom faktoren grammatikk og de standardiserte språktestene. For faktoren antonymer/synonymer og de standardiserte språktestene er korrelasjonene svakere, men tilfredsstillende. Det er derimot ikke samsvar mellom faktoren kategorisering og språktestene Ekspressivt Ordforråd fra CELF-4, Ordforståelse fra WPPSI-IV/WISC-IV og Bus Story. Kategorisering viser svake korrelasjoner med BPVS-II og TROG-2.

[Tabell 5 inn her]

Diskusjon

Hensikten med denne studien var å undersøke komponentene til kartleggingsverktøyet NSL gjennom analyser av reliabilitet, begrepsvaliditet og kriterivalidet. Det ble funnet tilfredsstillende reliabilitet og resultatene fra de konfirmerende faktoranalysene indikerte at NSL skiller mellom fire komponenter; kategorisering, vokabular, antonymer/synonymer og grammatikk, da firefaktormodellen viste den beste modelltilpasningen til dataene. I denne fire-faktormodellen ble det funnet invarians. Det operasjonaliserte begrepet språk ble målt på samme måte for kjønn og språkstatus, men ikke klassetrinn. Det ble i tillegg påvist signifikant positive korrelasjoner mellom NSL og standardiserte språktester (BPVS-II, Ekspressivt

Ordforråd fra CELF-4, Ordforståelse fra WPPSI-IV/WISC-IV, TROG-2 og Bus Story). Disse funnene belyser hvilke slutninger som kan trekkes av resultatene i lys av teori og empiri, og hvilken nytteverdi NSL kan ha for praksisfeltet.

Konfirmerende faktoranalyser sammenliknet en-faktormodellen (de ti delprøvene i NSL) med tre alternative flerfaktormodeller (to-faktormodell, tre-faktormodell eller fire-faktormodell), hvor fire-faktormodellen viste den beste modelltilpasningen. Det var signifikant høye korrelasjoner mellom komponentene kategorisering, vokabular, antonymer/synonymer og grammatikk i 2014-utvalget, noe som viser at komponentene er nært relaterte ferdigheter. Ut ifra disse korrelasjonene, kan det diskuteres hvor forskjellig komponentene faktisk er. Det var imidlertid lavere korrelasjoner mellom disse komponentene i 2016-utvalget. Særlig kategorisering viste svakere korrelasjoner med de tre øvrige komponentene, da spesielt med vokabular. En årsak til forskjellen i korrelasjonene mellom utvalgene kan være kriteriene for seleksjon. Mens 2014-utvalget bruker hele elevvariasjonen på klassetrinnene, har 2016-utvalget kun med språksvake elever. Variasjonen mellom barna i 2016-utvalget er dermed mindre enn i 2014-utvalget. Dette påvirker korrelasjonene mellom faktorene siden disse korrelasjonene baseres på variasjonen som finnes i utvalget.

Ved å benytte et utvalg bestående av både enspråklige og minoritetsspråklige elever (2014-utvalget) var det mulig å gjennomføre analyser av invarians i NSL for å se om språk måles på samme måte i begge gruppene. NSL differensierte ikke mellom språkstatus. Tilsvarende resultat ble funnet for kjønn (språk ble målt på sammen måte hos jenter og gutter), mens NSL ikke målte det samme begrepet over klassetrinn (alder). Sistnevnte er naturlig siden det stilles ulike krav til elevenes ferdigheter, kunnskap og læringsmål på hvert enkelt klassetrinn og begrepene som er en del av NSL bygger på det som læres på tvers av fagene norsk, matematikk, samfunnsfag, naturfag og KRLE på ulike trinn. Studier har samtidig påvist betydelige forskjeller i språkferdigheter mellom enspråklige og

minoritetsspråklige elever (Melby-Lervåg & Lervåg, 2014). I denne studien ble det funnet

signifikante forskjeller mellom enspråklige og minoritetsspråklige elever innenfor vokabular,

antonymer/synonymer og grammatikk, mens det ikke ble påvist tilsvarende forskjeller for

kategorisering. Kategorisering skiller seg fra de tre andre komponentene ved at det i tillegg til

språkferdigheter stilles krav til logisk resonneringsevne i oppgavene. Siden kategorisering

ikke differensierer mellom enspråklige og minoritetsspråklige elever, kan det diskuteres

hvorvidt det er hensiktsmessig å inkludere denne typen oppgaver i et kartleggingsverktøy

hvor formålet er å vurdere språkforskjeller mellom elevene. Samtidig har inkluderingen av

kategorisering i denne studien vist at forskjellene mellom enspråklige og minoritetsspråklige

elever først og fremst er knyttet til vokabular (bredde og dybde) og grammatikk.

Et annet viktig aspekt ved kvalitetsvurderinger er hvorvidt måleinstrumentet

samsvarer med andre tester som er ment å måle det samme teoretiske begrepet

(kriterievaliditet), i dette tilfellet språk. NSL viste tilfredsstillende korrelasjoner med øvrige

standardiserte språktester (BPVS-II, Ekspressivt Ordforråd fra CELF-4, Ordforståelse fra

WPPSI-IV/WISC-IV, TROG-2 og Bus Story) som ble foretatt i 2016-utvalget på tre av fire

faktorer, slik at kriterievaliditet er oppfylt.

Begrensninger

Det er begrensinger ved studien som bør belyses. 2014-utvalget ble ikke kartlagt med

supplerende språktester, og analyser av kriterievaliditet ble derfor kun gjennomført på 2016-

utvalget. Dette er et utvalg elever som befant seg i den ene ytterkanten av

normalfordelingskurven (1.5 SD eller mer under gjennomsnittet). Samtidig er det viktig å

påpeke at det ble funnet kriterievaliditet under piloteringen av NSL gjennom signifikant høye

korrelasjoner mellom NSL og TROG-2 og NSL og delprøven Ordforståelse fra WISC-IV.

Disse resultatene var basert på et normalfordelt utvalg.

KARTLEGGING AV SPRÅKFERDIGHETER

Oppsummerende refleksjoner og pedagogiske implikasjoner

Språkferdigheter er viktige byggesteiner for barns tilegnelse av kunnskap ved at det gjør det mulig å forstå hva et enkeltord betyr, informasjonen som formidles i en setning, innholdet i en tekst eller i samtaler. Et springende punkt innenfor dagens utdanningssystem handler om utvikling av kartleggingsverktøy for identifisering av barn som trenger språkstøtte på andrespråket (Kunnskapsdepartementet, 2019). Dersom elever med mangelfulle språkferdigheter ikke fanges opp tidlig, vil det kunne ha konsekvenser for læringsutbyttet, da videreutvikling av skolefaglige ferdigheter bygger på det språklige grunnlaget et barn har. For å gi et bedre opplæringstilbud som på sikt kan bidra til å redusere det skoleakademiske gapet mellom elevene, er det viktig å kartlegge hvilke språklige komponenter som bør gis eksplisitt opplæring i undervisningen.

En forutsetning for at kartleggingsverktøy fungerer som et redskap for å identifisere barn som er i behov av språkstøtte i skolen, er at slutningene som trekkes fra resultatene er valide. Denne studien viste at NSL har tilfredsstillende reliabilitet, begrepsvaliditet og kriterievaliditet og kan så langt sees som egnet til kartlegging av enspråklige og minoritetsspråklige elevers språkferdigheter på 1.-4.trinn. NSL viser at det er spesielt vokabular som skiller mellom enspråklige og minoritetsspråklige elever, tett etterfulgt av grammatikk og antonymer/synonymer. Det er imidlertid ingen forskjeller mellom elevgruppene når det gjelder kategorisering. Klassetrinnbaserte normer er nødvendig ved tolkning av resultatene. Kartleggingsverktøy som NSL kan derfor være et viktig supplement i opplæringen når det gjelder å finne ut hvilke språkferdigheter en elev behersker, hvilke eleven er på vei til å mestre og hvilke tiltak som bør iverksettes for å redusere det språklige gapet.

Referanser

Arnesen, A., Braeken, J., Ogden, T., & Melby-Lervåg, M. (2019). Assessing

children's social functioning and reading proficiency: a systematic review of the

quality of educational assessment instruments used in Norwegian elementary schools.

Scandinavian Journal of Educational Research, 63(3), 465-490.

doi:10.1080/00313831.2017.1420685

Brown, T.A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York:

Guilford Press.

Bishop, D.M. (2003). *Test for Reception of Grammar version 2 (Trog-2 manual)*

[Norwegian translating and norms: Lyster, S.-A.H. & Horn, E., 2009]. Stockholm,

Sverige: Pearson

Dickinson, D.K., & Tabors, P.O. (2001). *Beginning literacy with language: Young children

learning at home and school*. Baltimore, MD: Brookes.

Dickinson, D. K., & Porche, M. V. (2011). Relation between language experiences in

preschool classrooms and children's kindergarten and fourth-grade language and

reading abilities. *Child Development*, *82*(3), 870–886. doi:10.1111/j.1467-

8624.2011.01576.x

Dunn, L.M., Dunn, L.M., Whetton, C., & Burley, J. (1997). *British picture vocabulary scale

II.* Windsor, UK: National Foundation for Educational Research-Nelson.

Frøyen, W., Ahmadinia, S., Heller, M.C., & Skjåk, O. (2011). *Norsk som læringsspråk, NSL*.

Oslo, Norge: PPT Oslo, Utdanningsetaten.

Frøyen, W., Ahmadinia, S., Heller, M.C., & Skjåk, O., & Namvar, N. (2015). *Norsk som

læringsspråk redigert, NSL-R*. Oslo, Norge: PPT Oslo, Utdanningsetaten.

Hagen, Å.M., Melby-Lervåg, M., & Lervåg, A. (2017). Improving language comprehension in preschool children with language difficulties: A cluster randomized trial. *Journal of Child Psychology and Psychiatry, 58*(10), 1132-1140. doi:10.1111/jcpp.12762

Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday life of young American children*. Baltimore: Paul Brookes.

Hayiou-Thomas, M.E., Dale, P.S., & Plomin,,R. (2012). The etiology of variation in language skills changes with development: a longitudinal twin study of language from 2 to 12 years. *Developmental Science, 15*(2), 233-249. doi:10.1111/j.1467-7687.2011.01119.x

Heller, M.C., Lervåg, A., & Grøver, V. (2019). Oral language intervention in Norwegian schools serving young language-minority learners: a randomized trial. *Reading Research Quarterly, 54*(4), 531-552. doi:10.1002/rrq.248

Hjetland, H., Brinchmann, E.I., Scherer, R., & Melby-Lervåg, M. (2017). Preschool predictors of later reading comprehension ability: a systematic review. *Campbell Systematic Reviews, 13*(1). doi:10.4073/csr.2017.14

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, *26*(1), 55–88. doi:10.1016/j.dr.2005.11.002

Hu, L-T., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. doi:10.1080/10705519909540118

Klem, M., Melby-Lervåg, M., Hagtvet, B.E., Lyster, S.-A.H., Gustafsson, J.E., & Hulme, C. (2015). Sentence repetition is a measure of children's language skills rather than working memory limitations. *Developmental Science, 18*, 146-154.

Kline, R.B. (2016). *Principles and practice of structural equation modeling* (4th edition). New York: The Guilford Press.

Kunnskapsdepartementet (2019). Tett på – tidlig innsats og inkluderende fellesskap i barnehage, skole og SFO. Meld. St. 6 (2019-2020). Oslo: Departementets servicesenter.

Lervåg, A., Hulme, C., & Melby-Lervåg, M. (2018). Unpicking the developmental relationship between oral language skills and reading comprehension: it's simple but complex. *Child Development, 89*(5), 1821-1838. doi:10.1111/cdev.

Lyster, S-A.H., Horn, E. & Rygvold, A-L. (2010). Ordforråd og ordforrådsutvikling hos norske barn og unge: Resultater fra en utprøving av British Picture Vocabulary Scale, Second Edition (BPVS-II). *Spesialpedagogikk*, *9*, 35–43.

Melby-Lervåg, M., & Lervåg, A. (2014). Reading comprehension and its underlying components in second-language learners: a meta-analysis of studies comparing first- and second-language learners. *Psychological Bulletin, 140*(2), 409-433. doi:10.1037/a0033890

Melby-Lervåg, M., Lervåg, A., Lyster, S.-A.H., Klem, M., Hagtvet, B.E., & Hulme, C. (2012). Nonword repetition ability does not appear to be a causal influence on children's vocabulary development. *Psychological Science, 23*, 1092-1098. doi:10.1177/0956797612443833

Messick, S. (1995). Validation of inferences from persons' responses and performances as scientific inquiriy into score meaning. *American Psychologist, 50*(9), 741-749. doi:10.1037/0003-066x.50.9.741

Muthén, L.K., & Muthén, B.O. (2017). *Mplus user's guide* (8th ed.)*.* Los Angeles, CA: Muthén & Muthén.

Putnik, D.L., & Bornstein, M.H. (2016). Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. *Developmental Review, 41*, 71-90. doi:10.1016/j.dr.2016.06.004

Renfrew, C.E. (1997). *Bus story test- A test of narrative speech* (4th ed.). Bicester, England: Winslow Press.

Rogde, K., Melby-Lervåg, M., & Lervåg, A. (2016). Improving the general language skills of second-language learners in kindergarten: A randomized controlled trial. *Journal of Research on Educational Effectiveness, 9(1), 150-170. doi:10.1080/19345747.2016.1171935*

Rowe, M.L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development, 83(5), 1762-1774. doi:10.1111/j.1467-8624.2012.01805.x*

Semel, E., Wiig, E.H., & Secord, W.A. (2003). *Clinical evaluation of language fundamentals (4th ed.) [Translated and adapted to Norwegian]*. Toronto, Canada and *Assessio, Norway:* The Psychological Corporation/Pearson Assessment Company.

Snow, C.E., & Uccelli, P. (2009). The challenge of academic language. In D.R. Olson and N. Torrance (Eds), *The Cambridge handbook of literacy* (pp. 112-133). Cambridge: Cambridge University Press.

Statistisk sentralbyrå (2019). Elever med morsmålsopplæring, tospråkleg fagopplæring, tilrettelagd opplæring og særskild norskopplæring. https://www.ssb.no/utdanning/statistikker/utgrs

Storch, S.A., & Whitehurst, G.J. (2002). Oral language and code-related precursors to reading: evidence from a longitudinal structural model. *Developmental Psychology, 38*, 934-947. doi:10.1037/0012-1649.38.6.934

Thorndike, R.M., & Thorndike-Christ, T. (2014). Measurement and evaluation in psychology and education (8th edition), Harlow: Pearson.

Tomblin, J.B., & Zhang, X. (2006). The dimensionality of language ability in school-

age children. *Journal of Speech, Language, and Hearing Research, 49*(6), 1193-1208.

doi:10.1044/1092-4388(2006/086)

Utdanningsdirektoratet (2006). Læreplanverket Kunnskapsløftet (LK06). Oslo:

Utdanningsdirektoratet.

Wechsler, D. (2003). *Wechsler intelligence scale for children administration manual*

(4th ed.) [Norwegian administration manual]. San Antonio, TX and Stockholm,

Sweden: The Psychological Corporation/Pearson Assessment Company.

Wechsler, D. (2012). *Wechsler preschool and primary scale of intelligence administration*

manual (4th ed.) [Norwegian administration manual]. San Antonio, TX and

Stockholm, Sweden: The Psychological Corporation/Pearson Assessment Company.

Tabell 1

Gjennomsnitt, standardavvik og reliabilitet for 2014-utvalget og 2016-utvalget

	2014-utvalget			**2016-utvalget**		
	Gjennomsnitt	Standardavvik	Cronbachs α	Gjennomsnitt	Standardavvik	Cronbachs α
Enfaktor modell	129.99	23.55	.959	64.01	19.80	.939
Firefaktor modell						
Kategorisering	42.57	10.04	.922	24.87	9.47	.911
Vokabular	47.72	8.50	.899	19.92	10.50	.934
Antonymer/synonymer	11.09	6.23	.898	1.68	2.15	.730
Grammatikk	28.62	3.12	.751	17.45	6.45	.877

Tabell 2

Korrelasjoner mellom delprøver for 2014-utvalget (horisontal triangel) og 2016-utvalget (vertikal triangel)

	1	2	3	4	5	6	7	8	9	10
1. Matching av kategorier (MK)	1	.195*	.003	.008	.138	.175*	.067	.054	.076	.140
2. Generalisering av kategorier (GK)	.391**	1	.080	.044	.085	.169*	.157	.266**	.148	.214*
3. Ordforståelse impressiv (OFI)	.448**	.356**	1	.489**	.191*	.121	.452**	.264**	.508**	.394**
4. Ordforståelse ekspressiv (OFE)	.446**	.355**	.517**	1	.263**	.420**	.514**	.395**	.534**	.473**
5. Antonymer (ANT)	.538**	.395**	.585**	.624**	1	.385**	.237**	.154	.244**	.172*
6. Synonymer (SYN)	.420**	.237**	.369**	.418**	.553**	1	.275**	.247**	.329**	.293**
7. Adjektiv (ADJ)	.484**	.401**	.545**	.530**	.574**	.381**	1	.624**	.610**	.562**
8. Preposisjoner (PRE)	.481**	.361**	.563**	.503**	.537**	.333**	.575**	1	.604**	.605**
9. Pronomen (PRO)	.422**	.369**	.402**	.325**	.417**	.282**	.426**	.412**	1	.477**
10. Adverb (ADV)	.352**	.300**	.419**	.436**	.464**	.340**	.447**	.433**	.282**	1

Merknad: * korrelasjoner er signifikant på .05 nivå, ** korrelasjoner er signifikant på .01 nivå

Tabell 3

Faktormodeller for 2014-utvalget

	χ^2 (*df*), *p*	CFI	TLI	RMSEA	SRMR	Adj. BIC
Faktormodell 1	65.868 (35), .001	.979	.972	.049	.030	15785
Faktormodell 2	63.283 (34), .002	.980	.973	.048	.029	15786
Faktormodell 3	50.048 (32), .022	.987	.982	.039	.027	15778
Faktormodell 4	31.520 (29), .341	.998	.997	.015	.023	15768

Tabell 4

Faktormodeller for 2016-utvalget

|	χ^2 (*df*), *p*	**CFI**	**TLI**	**RMSEA**	**SRMR**	**Adj. BIC**
Faktormodell 1	82.871 (35), .000	.880	.845	.100	.066	6264
Faktormodell 2	78.556 (34), .000	.888	.852	.098	.063	6261
Faktormodell 3	64.347 (32), .001	.919	.886	.086	.058	6251
Faktormodell 4	48.213 (29), .014	.952	.925	.070	.044	6240

Tabell 5

Korrelasjoner mellom komponentene i firefaktormodellen i Norsk Som Læringsspråk (2016-utvalget) og de standardiserte språktestene

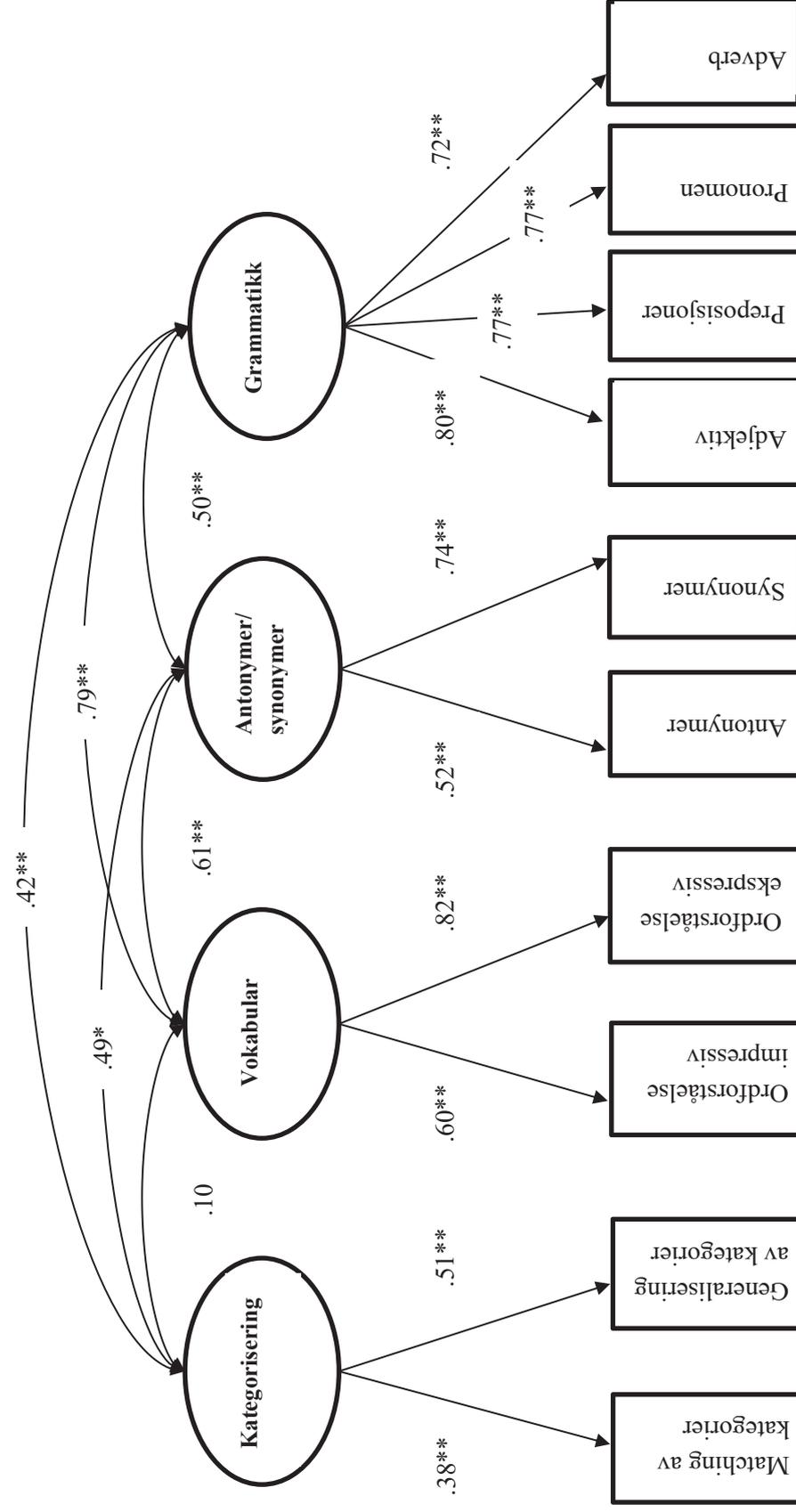|	1	2	3	4	5	6	7	8	9
1. British Picture Vocabulary Scale II (BPVS-II)	1								
2. Ekspressivt Ordforråd (Clinical Evaluation of Language Fundamentals, CELF-4)	.557**	1							
3. Ordforståelse (WPPSI-IV/WISC-IV)	.541**	.487**	1						
4. Test for Reception of Grammar (TROG-2)	.643**	.372**	.535**	1					
5. Bus Story	.494**	.522**	.563**	.457 **	1				
6. Kategorisering	.228**	.004	.007	.303**	.065	1			
7. Vokabular	.531**	.395**	.632**	.537**	.437**	.035	1		
8. Antonymer/synonymer	.260**	.211*	.407**	.413**	.302**	.206*	.322**	1	
9. Grammatikk	.545**	.359**	.589**	.621**	.473**	.168*	.617**	.336**	1

Merknad: * korrelasjoner er signifikant på .05 nivå, ** korrelasjoner er signifikant på .01 nivå

Figur 1. Firefaktormodell for 2014-utvalget.
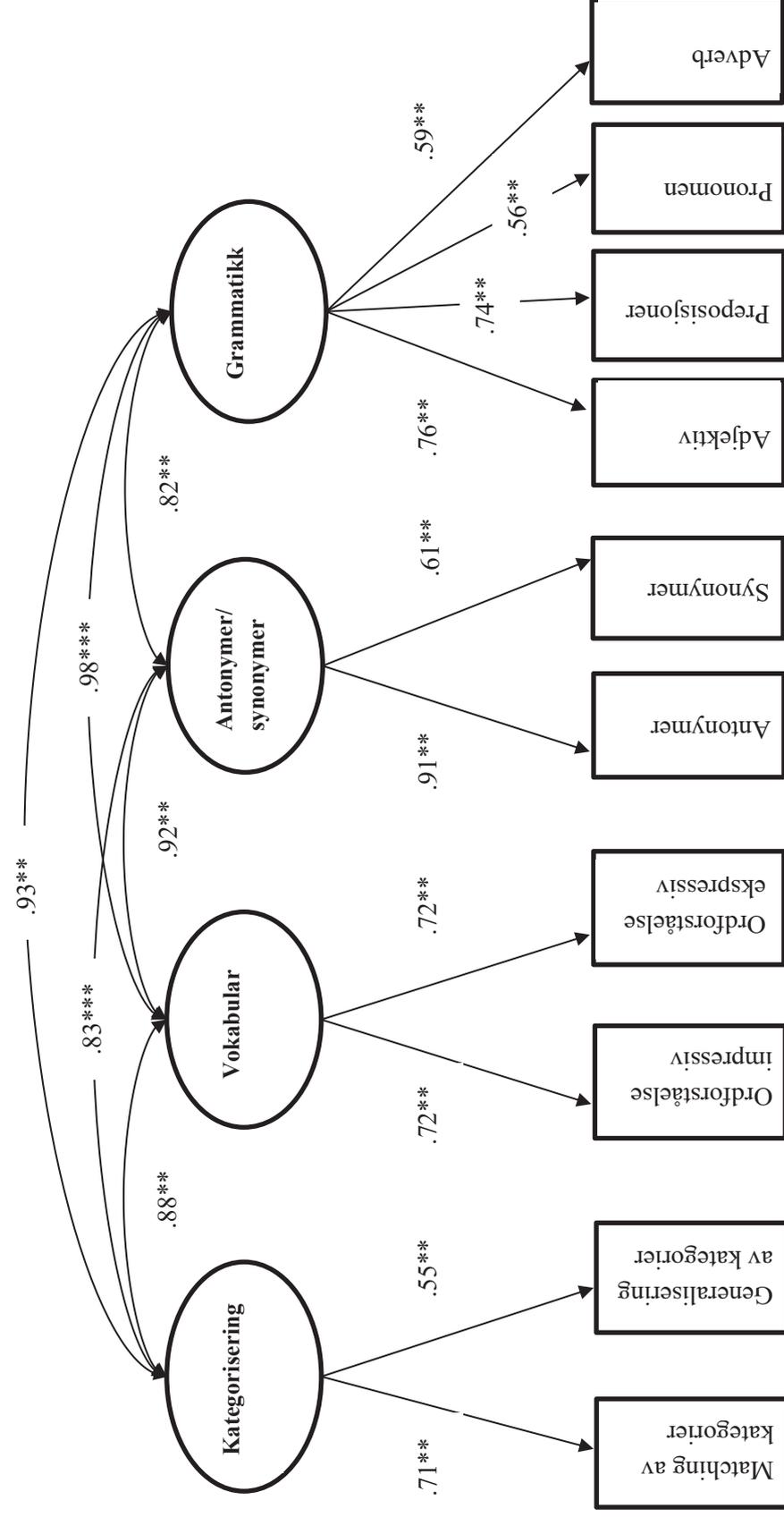
Merknad: * korrelasjoner er signifikant på .05 nivå, ** korrelasjoner er signifikant på .01 nivå.

Figur 2. Firefaktormodell for 2016-utvalget.

Merknad: * korrelasjoner er signifikant på .05 nivå, ** korrelasjoner er signifikant på .01 nivå.