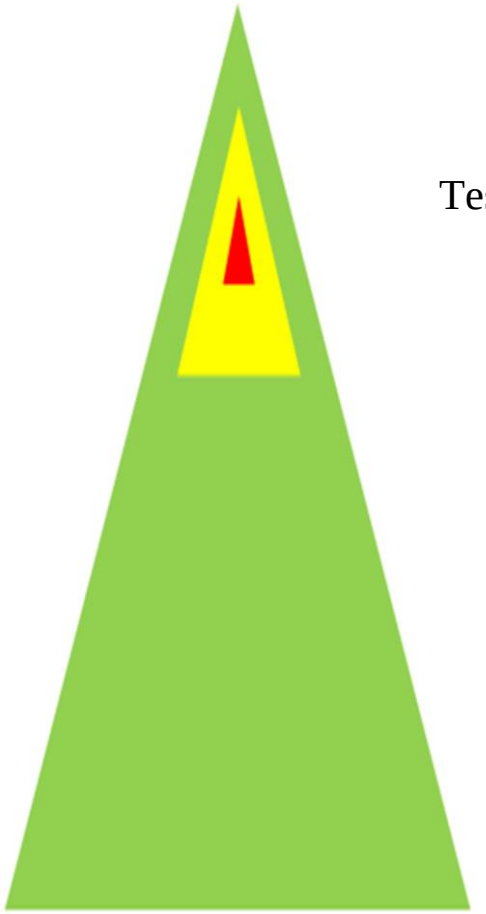


Testing the effectiveness of the N-PALS model

– a school-wide framework to prevent externalizing student problem behavior

Mari-Anne Sørlie

Dissertation (dr. philos) submitted to the Faculty of Educational Sciences, University of Oslo.

Series of dissertations submitted to the
Faculty of Educational Sciences, University of Oslo
No. 333

ISSN 1501-8962

Cover: Hanne Baadsgaard Utigard.
Print production: Reprosentralen, University of Oslo.

Summary

In this this dissertation, school level impacts of the school-wide prevention model SWPBS (School-Wide Positive Behavior and Support), as implemented in a sample of Norwegian elementary schools, were investigated after three years of implementation. SWPBS is known under the acronym PALS in Norway (N-PALS in English). N-PALS is a three-tiered (universal, selected, indicated) inclusive system approach which schools can use to prevent and reduce externalizing student behaviors, and to establish a positive and supporting learning climate for all students. Short-term impacts of an abbreviated version of N-PALS (PPBS, Preventing Problem Behavior in School) were also investigated. Additionally, an underlying assumption in N-PALS, that differences between schools in prevalence of student problem behavior are related to psychosocial characteristics of the school context, was addressed.

Intervention effects on moderate and severe problem behavior in and outside the classroom context, on classroom climate and inclusion, and on perceived collective efficacy, self-efficacy, and disciplinary practices among the school staff were examined using a strengthened nonequivalent control group (NEC) design and multi-level analyses. Differential intervention effects were examined with implementation quality (fidelity), school size, program training dosage, or portion of untrained staff as moderators. Selected data from NUBU's (Norwegian Center for Child Development) longitudinal effectiveness and development study 'Positive Behavior Support in School' were used.

This study involved totally 12,050 students (Grades 4-7), 3,695 teachers, after school personnel, and principals, and a random sample of 2,750 parents from 65 elementary schools around the country. Twenty-eight of the schools implemented the full-scale N-PALS model, 17 schools implemented the abbreviated version (PPBS), and 20 schools doing 'practice-as-usual' (PAU) made up the control group. The study was conducted over five successive

school years. The analyses and results included in the dissertation are based on staff and student ratings from five of the totally six assessment points (double pre-test, two intermediate tests, post-test), and cover four of the five study years.

The N-PALS model seemed effective in reducing the level of both moderate and severe externalizing student problem behaviors (particularly on common school areas) and in promoting qualitatively better social and academic learning conditions in class. The results also indicate increased ability in the intervention schools to reach out to all students and to reduce the number of students singled out for individual intervention plans due to conduct problems. N-PALS also appeared effective in promoting the school staffs' perceived individual and collective efficacy. Besides, this three-level prevention framework appeared effective in changing staff behavior in positive ways. The schools with high implementation quality benefitted the most from the intervention model, and small-moderately sized schools reached higher implementation scores than large schools.

Outcome analyses of the abbreviated PPBS intervention indicated promising immediate effects on several but not on all of the same outcome variables as the full three-level model. Moderation analyses indicated better outcomes for schools with high fidelity scores, high mean training dosage, higher portion of untrained staff, and for small to medium sized schools. The significant effect sizes for the two intervention frameworks were generally large enough to be considered of practical significance, and the confidence intervals were narrow, indicating high precision of the estimates. The study results indicate that SWPBS was successfully transferred across the Atlantic without major adaptations, without violating the national curriculum or the norm of inclusion.

The current study adds to the existing base of knowledge in several ways, including by being the first to examine the cumulative effects of the three-level model and by being the first effectiveness study of the SWPBS model implemented in Europe and in a non-English

speaking context. Moreover, the Norwegian study is the first investigating effects of SWPBS/N-PALS on perceived teacher collective efficacy and on school staffs' disciplinary practices, and also one of very few longitudinal effectiveness studies to date that have been conducted on this school-wide prevention model. Besides, our study is the first to investigate the relationship between collective efficacy and problem behavior in school, and on an empirical basis suggesting that teacher collective efficacy is a significant and reliable indicator of a school's culture. Ideally, the study results should be replicated in a Scandinavian randomized controlled trial before more firm conclusions are drawn on the effectiveness of the N-PALS model.

Acknowledgements

The intervention study on which this work builds, was initiated and implemented by the Norwegian Center for Child Behavioral Development (NUBU), and partially financed by contributions from the Norwegian Directorate for Education and Training.

I am grateful to NUBU for the opportunity to being the main researcher of this challenging large-scale study, and for the opportunity to finalize a doctoral thesis. In particular, I would like to thank Terje Ogden, Asgeir Røyrhus Olseth, and Torbjørn Torsheim for your crucial contributions to this work – and not at least, for your patience and unconditional positive support over many years. Without your guidance, critical inputs, willingness to share your comprehensive expertise, and analytical skills, this work would have run aground.

Not at least – a big thank to the 65 schools, and each student and school staff member contributing to the study. Without your willingness to share your time, experiences, and opinions in the completion of our questionnaires - again and again - the study would have been impossible. I hope you find the study results interesting and practically beneficial.

List of papers

1. Sørlie, M-A. & Ogden, T. (2014). Reducing threats to validity by design in a nonrandomized experiment of a school-wide prevention model. *International Journal of School & Educational Psychology, 2,* 235-246. Doi:10.1080/21683603.2014.881309

2. Sørlie, M-A. & Ogden, T. (2015). School-Wide Positive Behavior Support-Norway: impacts on Problem behavior and classroom climate. *International Journal of School & Educational Psychology, 00,* 1-16. Doi: 10.1080/21683603.2015.1060912

3. Sørlie, M-A., Ogden, T., & Olseth, A. R. (2016). Examining teacher outcomes of the School-Wide Positive Behavior Support Model in Norway: Perceived efficacy and behavior management. *Sage Open, April-June 2016,* 1-13. Doi: 10.1177/2158244016651914

4. Sørlie, M-A., Ogden, T., & Olseth, A. R. (2015). Preventing problem behavior in school through school-wide staff empowerment: Intervention outcomes. *World Journal of Educational Research, 2*, 117-139.

5. Sørlie, M-A. & Torsheim, T. (2011). Multilevel analysis of the relationship between teacher collective efficacy and problem behavior in school. *School Effectiveness and School Improvement, 2*, 175-191. Doi: 10.1080/09243453.2011.563074

Contents

List of Figures

List of Tables

1. Introduction

1.1 Focus of the Dissertation

The main focus of this dissertation is on evaluating the effectiveness of the *PALS model* (Norwegian acronym for positive behavior, learning environments and interactions in school; called N-PALS in English) and an abbreviated version of the model (PPBS, preventing problem behavior in school). Considerable efforts were made to produce as valid effect estimates as possible using a strengthened quasi-experimental design. Our concerns and efforts made in the design phase of the study are highlighted and tested in Paper 1. In Paper 2 the relative school level effects of the N-PALS model on incidents of problem behavior within and outside the classroom context and on the classroom climate are investigated. Paper 3 focuses on intervention benefits for the school staff. Paper 4 concentrates on outcomes of the PPBS intervention. The outcome papers (2-4) additionally investigate if some schools had better outcomes than others, and if so, under which conditions were the outcomes better and worse? In Paper 5, a vital underlying assumption in N-PALS is emphasized; that is, how students behave in school is significantly influenced by the school itself.

N-PALS is a three-tiered inclusive system approach which schools can use to prevent and reduce externalizing student behaviors, and to establish a positive and supporting learning climate for all students. Systematic and collective use of positive and proactive supports, and other evidence-based practices are at the core of N-PALS. This school-wide intervention model is a culturally adapted version of the SWPBS model (School-Wide Positive Behavior and Support), developed in the USA. A variety of positive impacts have been reported across studies (Horner, Sugai, & Lewis, 2015), of which the large gross have been conducted in the USA. The SWPBS model was adapted and introduced to the Norwegian school system by special advisor Anne Arnesen and Wilhelm Meek-Hansen at the Norwegian Center for Child

Behavioral Development (NUBU), and named N-PALS. Later, they developed an abbreviated version, called the PPBS intervention (preventing problem behavior in school).

In 2002, the implementation of N-PALS started out with four primary schools. Today, 230 schools are implementing the N-PALS model, which include 9.7% of all primary (Grades 1-7, N = 1.615) and combined schools (Grades 1-10, N = 746) in Norway (SSB, 2017). Since the start-up, 80 professionals have been trained as N-PALS supervisors (54 still active). Around the world, nearly 26.000 elementary schools, junior high schools, and high schools (most in the USA) are implementing SWPBS (www.pbis.com). It follows, that SWPBS is one of the most popular and far-reaching preventive interventions that exist per date.

The effectiveness of N-PALS in Norway was first evaluated in a pilot study with promising results after two years of implementation (Sørlie & Ogden, 2007; Ogden, Sørlie, & Hagen, 2007). However, because only four intervention schools and four neighboring comparison schools participated, the validity of the results could be questioned. In 2007, a second effectiveness study with a more robust design was initiated. The candidate has been the main researcher of this longitudinal large-scale study, in which 65 elementary schools (Grades 1-7) around the country participated. Twenty-eight of the schools were implementing the N-PALS model, 17 schools were implementing the PPBS intervention, while the control group comprised of 20 schools doing “practice-as-usual”. More than 12.000 students (grades 4-7), 3.600 staff members (teachers, assistants, special education teachers, after-school personnel, principals), and 2.750 parents (random sample) contributed with data at six assessment points across the five-year study period. Analyses of some of these data make up the empirical basis of the dissertation. The comprehensive amount of data is still being analyzed, and new results are successively being published.

In addition to the candidate (full-time), the research staff comprised of Asgeir Røyrhus Olseth (research coordinator, co-author, full-time), Terje Ogden (prior research director at

NUBU, formally responsible for the study, co-author), and Torbjørn Torsheim (professor at the University of Bergen, 20% part-time post at NUBU for one year, analytical support, co-author). None of the model purveyors or adaptors were involved in the design or conduct of the study, nor in the analytical and publication processes.

In the following, the "back cloth" of the study and the political and professional rationales for introducing N-PALS to Norwegian schools are described. The situation in Norwegian schools in the years prior to the introduction of this preventive framework was characterized by major challenges related both to student problem behavior and inclusive education. Further on in this chapter, emphasizes are put on how behavior problems in children and youth can be defined and categorized, how large the problem of student misconduct seems to be in Norway, potential risk and protective factors related to student problem behavior, and on how problem behavior often is handled in school. Finally, characteristics of effective school-based interventions are highlighted.

1.2 Effective Prevention and Inclusive Education: Key Challenges in Norwegian Schools

There is an extensive body of research substantiating that the school is an influential and even critical developmental context for children and youth, not only for their academic learning but also for their conduct and psycho-social well-being (e.g., Gage, Larson, Sugai, & Chafouleas, 2016; Gottfredsson et al., 2005; Korpershoek, Harms, De Boer, van Kuijk, & Doolaard, 2016; Pauwels & Svensson, 2015; Thapa, Cohen, Guffey, & D'Alessandro, 2013; Wang & Degol, 2016; Wissink et al., 2014). There are several arguments for schools being excellent arenas for prevention of behavioral and emotional problems among children and youths: a) schools have better access to young people and their parents than any other social institution, b) interventions can be offered within ordinary childhood environments and under secure conditions, c) interventions can be universal or offered to vulnerable groups and individual

children at high risk without segregation and stigmatizing effects, and d) teachers hold significant competence and interact with the children on a daily basis over time (e.g., Arnesen, Ogden, Sørlie 2006; Helgeland 2008; Ogden 2009).

In line with this, the Norwegian Education Act (§9a), the national curriculum and specified principles of teaching, emphasize that Norwegian schools are obliged to give all children and youths a safe and orderly school environment in which both student learning and positive social development should be the governing objectives - a policy consistent with the UN Convention on the Rights of Children (UNICEF, 1989) and the Salamanca Statement (UNESCO, 1994). Education should be inclusive, that is, all students, regardless of any challenges they may have, should be placed in age-appropriate general education classes in their neighborhood school to receive high-quality instruction, interventions, and supports that enable them to feel respected, confident, and safe so he or she can learn and develop to his or her full potential (Alquraini & Gut, 2012). Accordingly, inclusive education is not just a matter of promoting the students' academic skills, but also a matter of stimulating and supporting the students' personal development and identity.

Despite the governing goals of inclusive schools and adaptive education for more than 40 years, effective prevention of school misconduct and inclusive ways of handling students with behavioral difficulties still represent a demanding task for Norwegian schools, as for schools in our neighboring countries Finland, Sweden and Denmark (e.g., Dyssegaard, Larsen, & Tiftikci, 2013; EVA, 2011; Kimber, 2011; Markussen, 2002; OECD, 2014). Behavior problems among students have long been a concern for teachers, school-leaders, and policy makers, and has been described as one of the largest unmet challenges in our school system (e.g., Nordahl, Sørlie, Manger, & Tveit, 2005; Ogden, Sørlie, Arnesen, & Meek-Hansen, 2012; Report to the Storting no. 30, 2003-2004).

Student problem behavior reduces the quality of teaching and the learning

environment, hampers positive social relations with classmates and teachers, and reduces the time to teach, which in turn may negatively affect the students social and academic functioning, sense of well-being, safety, and attachment to school (e.g., Houts, Caspi, Pianta, Arseneault, & Moffitt, 2010). High prevalence of problem behavior among students is said to be an indicator of schools that fail, and a correlate of reduced learning engagement and low school level achievements (Algozzine, Daunic, & Smith, 2010). At the individual level, problem behavior in childhood and adolescence has repeatedly across studies shown negative associations with a wide range of concurrent and longitudinal outcomes, such as low social competence, academic failure, and criminal offences (e.g., Broidy et al., 2003; Fergusson, Horwood, & Ridder, 2007; Heiervang et al., 2007; Loeber, Farrington, Stouthamer-Loeber, Moffitt, & Caspi, 1998; Ogders, Moffitt, & Broadbent, 2008, Reef, Diamantopoulou, van Meurs, Verhulst, & van der Ende, 2011; Sørlie, Hagen, & Ogden, 2008). Especially children showing high levels of externalizing problem behaviors are at increased risk of escalating conduct problems and negative life outcomes. International and national longitudinal studies demonstrate that these children are at high risk of academic problems and school drop-out, social skills deficits, drug abuse, antisocial personality disorder, criminality, mental health problems as well as increased morbidity and labor market problems (e.g., Fergusson et al., 2007; Lahey et al., 2006; Loeber, & Farrington, 2000; McCart, Ogden, & Henggeler, 2014; Sørlie et al., 2008).

Accordingly, there are good reasons to focus on what can be done in school to prevent and reduce problem behavior among children and youth in inclusive ways, and to identify school interventions and factors that are particularly influential on the students' behavior and learning conditions in school. Prevention implies early efforts to inhibit a negative behavioral or psychosocial development to start or worsen. Preventive efforts include initiatives and interventions to promote positive (prosocial) behavior and/or implementation of early

interventions to limit or eliminate contextual conditions and individual factors contributing (directly or indirectly) to student problem behavior, simultaneously contributing to the prevention of further escalation or relapse. Prevention of behavior problems in school thus is both a matter of "irrigation" (promote positive behavioral development and a favorable learning environment) and a matter of "weeding" (eliminate or reduce problem behavior and an unfavorable learning environment). Prevention represents the opposite of a "wait-and-see" approach, implying proactive and supporting measures in preference to reactive and punitive measures (Arnesen & Sørlie, 2012).

1.2.1 Growing concerns about student problem behavior. During the 1990's and the early 20's notes in the daily press and academic publications mirrored a growing concern about escalating behavior problems and increasing numbers of special needs students and segregated learning accommodations. One of our most respected researchers, Terje Ogden, expressed that "Behavior problems in school remain as perhaps one of the largest challenges in the 1990's" (Ogden 1990, p. 81). National and international studies confirmed that student problem behavior was a major challenge in Norwegian schools, and more so than in many other countries (e.g., Ogden, 1998; Kjærnsli, Lie, Olsen, Roe, & Turmo, 2004). Ogden's (1998) study of a representative sample of Norwegian schools showed that during a random week, more than half of the teachers experienced disturbing behavior that interferred with instruction and learning activities, while 10% reported of serious norm and rule breaking incidents. In 2000, OECD's first worldwide PISA-study (Programme for International Student Assessment) was performed, and showed that 39% of 15 year old students reported noise and disorder during most or all lessons. In 2004, the proportion was 40%. These results placed Norway at the bottom among the OECD-countries (Kjærnsli et al., 2004).

The concerns and empirical findings were followed by a growing political and

professional demand for more preventive, inclusive, and effective approaches to reduce and handle student problem behavior (e.g., KUF/BFD, 2000; Sørlie, 2000; Report to the Storting no. 28, 1998-1999, Report to the Storting no. 17, 1999-2000, Report to the Storting no. 30, 2003-3003) – a need that still applies (e.g., OECD, 2014; Official Report no. 2, 2015). For example, Report to the Storting no. 28 (1998-1999) identified prevention and handling of negative behavior as one of four priority areas. Report to the Storting no. 17 (1999-2000) emphasized the need to intensify, develop and coordinate the efforts to prevent and combat criminality among children and youths. As a consequence, the Norwegian Center for Child Behavioral Development (NUBU) was established with a nationwide responsibility to develop, evaluate and implement evidence-based programs for children and youth with or at high risk of developing conduct problems.

Moreover, Report to the Storting no. 17 (1999-2000) underlined the school's unique position and abilities to contribute both to general prevention and supports to vulnerable groups, and suggested a research-based evaluation of the efforts and programs used in Norwegian schools to prevent and handle behavior problems. Due to a general lack of knowledge on the effects of school-based programs, the Research Council of Norway commissioned and financed a review of international school intervention research targeting effects on serious behavior problems in school children. The review summarized characteristics of the most promising programs and concluded that school-based prevention programs generally have shown positive albeit small to moderate effects on student problem behaviors (Sørlie, 2000).

Later the same year, a report from a group of researchers was published on the use of various projects and programs to prevent behavior problems in Norwegian schools (KUF/BFD, 2000). It was summarized that reduction of problem behavior and promotion of prosocial skills were topics of great interest and efforts in many schools, but that the quality

was questionable, and evaluations were generally missing or of low validity. Only the Olweus' anti-bullying program (e.g., Olweus, 1991) was recommended for further implementation without the need for additional evaluation (KUF/BFD, 2000).

In 2002 the Government initiated the first of several successive "*Manifesto towards bullying*". Moreover, and in line with the disheartening facts and reviews of national and international research, two research-supported guidelines for how to handle severe behavior problems in school (commissioned by the Norwegian Directorate of Education and Training) were published in 2003, one for teachers and one for school administrators (Nordahl, Sørlie, Tveit, & Manger, 2003 a, b). Besides, a second report on the use of various school-based programs to prevent problem behavior and drug use was published in 2006 (Nordahl, Gravrok, Knutsmoen, Larsen, & Rørnes, 2006). The researchers concluded that considerable efforts are made in many schools to promote and reduce various student problems, but the initiatives are often reactive and weakly founded in theory and empirical data. It was added;

> "Besides, it is still the case that much of the work and the programs being implemented seldom are satisfactorily evaluated. This implies that much of the proactive and reactive practice in Norwegian schools strongly rest on subjective experiences, opinions in an area one knows little about, and on personal perceptions. Accordingly, we risk that much of the preventive efforts do not bring about desired results, and there is also a risk that some measures can be harmful for children and youths" (Nordahl et al., 2006, p. 155).

The growing political and professional concerns about student problem behavior by the turn of the century, and reports on the lack of evidence-based prevention in Norwegian schools, inspired the introduction and evaluation of the N-PALS model.

1.2.2 Increased focus on inclusive and successful education. In addition to escalating concerns about the level of problem behavior in our schools, the last two decades prior to the introduction of N-PALS were characterized by several school reforms and new nation-wide curricula, based respectively on a growing broad political consensus on the ideas of inclusive,

local, and differentiated adaptive education, and on worrying reports on the academic level in Norwegian schools. The ideas of inclusive schools and adaptive education were introduced in the 1970's, inspired by international ideological principles of integration, normalization and decentralization (Ogden, 2014), as expressed for example in the Salamanca Statement (UNESCO, 1994):

> "Children with special needs should have access to regular schools capable of meeting their needs, and regular schools with an inclusive orientation are the most effective means of building an inclusive society and achieving an education for all".

Accordingly, in 1992 all state special boarding schools for behavior problematic and mentally retarded children were closed down or restructured into regional "Statped"-units (Norwegian Support System for Special Needs Education) with services covering all types of disabilities. All students should now be included in ordinary schools, independent of their behavior, physical and cognitive abilities. The changes in the educational policy caused considerable challenges for ordinary schools. For example, a research report commissioned by the Norwegian Directorate of Education and Training elucidated that inclusion of behavior problematic students in ordinary schools seemed especially challenging (Jahnsen, Nergaard, & Flaatten, 2006). The authors concluded on the basis of a nationwide survey that the number of small-group measures established outside of ordinary school as a way of handling misbehaving students almost tripled in the period 1991-2005, and the number of students in such segregated educational settings more than doubled in 15 years.

However, results from international comparative studies published at the beginning of the new millennium moved the focus away from inclusion and behavior problems towards academic performance. Results across studies, such as the OECD's PISA-studies TIMSS (Trends in International Mathematics and Science Study) and PIRLS (Progress in International Reading Literacy Study) pointed to that the academic skills of Norwegian students were at or below the mean of the more than 30-60 participating countries

(Utdannings- og forskningsdepartementet, 2004). The results were surprising and provoked a shock-based political consensus on ameliorating the students learning outcomes. In 2006 a resource demanding school-reform were initiated, the Knowledge Promotion Reform, which implied new nationwide curriculum and a new quality assurance system.

Taken together; except for an increased political and professional focus on inclusion and bullying during the late 90's and early 20's, the schools' need for more knowledge and effective preventing practices towards the broader spectrum of student problem behavior, were not met. This omission contrasts the parallel challenges related to the students' academic achievements, in that no equivalent national inputs have been implemented, such as changes in the curriculum and teacher training or in increased resources to school-based programs to prevent student problem behaviors. As a contributing effort to meet the schools obvious need for more adequate, inclusive and effective ways to prevent and reduce the level of student problem behavior, implementation and evaluation of the N-PALS model was initiated by NUBU in 2002. Eighteen years later, this innovative work and support is still relevant and continuing.

1.3 Defining Behavior Problems in School

Behavior problems can be more or less serious, dependent on the severity, intensity, frequency, chronicity and consequences of the behaviors.

1.3.1 Clinical definitions (diagnoses). Within the psychological and psychiatric fields (psychopathology), assessment and criteria-based distinctions are made between problems and disorders. As concerns *externalizing disorders*, the 'Diagnostic and statistical manual of mental disorders' (APA, 2013) divides severe behavior problems into several sub-categories or diagnoses such as Oppositional Defiant Disorder (ODD) and Conduct Disorder (CD). ODD

involves a combination of angry or irritable mood, argumentative or defiant behavior, and vindictiveness or spitefulness shown at least twice within the last ½ year. The child's behavior relates to his or her own distress or that of others close to him or her, or it has a negative impact on the person's ability to function. CD is present if a person under 18 years and over time demonstrates a constellation of symptoms, including a combination of aggressive behavior toward people or animals and destruction of property, lying, theft, and serious violations of normally accepted rules of conduct. The person's behavior seriously and negatively affects his or her well-being, daily functioning and doings, learning and social interactions (APA, 2013). A person with mental or behavior *problems* shows symptoms that substantially influences his or her well-being, daily functioning and doings, learning and social interactions, but without the criteria of a diagnosis is satisfied.

Research has revealed that the development of conduct disorder/problems may follow different pathways or trajectories. Moffitt's dual taxonomy model (1993) suggests that different developmental courses define children's onset of severe behavior problems; an adolescent-limited (AL) or late-onset trajectory (i.e., onset in adolescence with desistance in adulthood) and a life-course persistent (LCP) or early-onset trajectory (i.e., onset in childhood with little desistance over time). According to this model, children on an early-onset trajectory are more vulnerable and are exposed to more risk factors on multiple domains than children on a late-onset trajectory, including neuropsychological dysfunction, mental-health problems, poor parenting, substance use disorders, learning difficulties, poor school attendance, head injuries, and childhood maltreatment and trauma (e.g., Kroll, et al., 2002, Miller, 2002). By contrast, AL is more common, transient, and near-normative (Silberg et al., 2015, p. 826) and probably has its origins in a 'maturity gap' and social processes; especially influenced by antisocial peer group pressures. Comparted with the LCP, AL is considered less heritable and less associated with neurodevelopmental features. This trajectory begins in adolescence but

desists in early adulthood.

Fairchild, Van Goozen, Stollery, & Goodyer (2008) reported, however, that both CD subtypes display emotion-processing deficits, changes in the brain structure and function, and alteration in the cortisol secretion. Johnson, Kemp, Heard, Lennings, & Hickie (2015) found that both groups had severely impaired executive functioning, but that the early-onset group additionally displayed deficits in verbal learning and memory, had higher rates of psychosis, childhood maltreatment and showed more serious violent behavior. LCP probably has its origins in neurodevelopmental deficits and/or hyperactivity but is exacerbated by high-risk social environments (e.g., inadequate parenting, family dysfunction, parental depression, poverty) (e.g., Silberg, Moore, & Rutter, 2015). Antisocial behavior is evident in childhood but worsens during adolescence. Across studies, this pattern or trajectory is found to be rather uncommon but pathological, although less persistent than suggested by Moffitt in that about 50% do not continue their antisocial behavior into adulthood (Ogders et al., 2008).

Although a diagnostic approach to behavior problems in school is not used in the current study, variability (heterogeneity) in the expression, frequency and severity of behavior problems in Norwegian schools should be expected (e.g., a small group at high risk, some at moderate risk and most at low risk of CD) and addressed when considering preventive provisions, for example, as in the N-PALS model.

1.3.2 Psychometrically based definitions. Researchers employing statistical approaches to child behavior problems have derived from multiple studies and sources (e.g., parent and teacher ratings, case records, observations) that problem behavior among children and youth can be dichotomized into two broad-band categories or syndromes (not equivalent to diagnoses): *externalizing* (extrovert) and *internalizing* (introvert) problems. In everyday language, it is the first category which is referred to when problem behavior among school

children is the topic. Externalizing problems can be described as outward-directed, non-compliant, aggressive, hostile or violent behavior. Externalizing behavior often generates discomfort for others and/or conflict in the environment. In contrast, internalizing problems can be described as inward-directed and primarily generating distress in the individual, including symptoms such as sadness, unhappiness, loneliness, depression, nervousness, and anxiety (Forns, Abad, & Kirchner, 2014). With reference to Sørlie and Nordahl's (1998) study of behavior problems in a sample of Norwegian primary, secondary, and upper secondary schools, externalizing problem behavior in school include 'classroom disruption' (i.e., behavior incompatible with instruction and learning), 'out-acting' (i.e. verbal and physical aggression) and 'antisocial behavior' (i.e., serious norm and rule breaking behavior). Internalizing problems correspond to 'social isolation' and is sometimes referred to as emotional problems in which anxiety and depression make up the core. The stability over time for both externalizing and internalizing problems is found to be in the moderate to high range, while some increases in the prevalence have been reported during adolescence (e.g., Hamdi & Iacono, 2014). Although internalizing and externalizing problems may co-occur (comorbidity), the dissertation's emphasis is on externalizing problem behavior.

1.3.3 School problem behavior as defined in the dissertation. In the educational field, no commonly accepted definition exists that takes into account the severity of the students' problem behavior or its relative, contextual, and transactional features. Behavior problems are usually defined based on the teachers' individual perceptions and judgements of the students' behavior, and spoken of as 'social and emotional problems', 'emotional and behavioral problems' or 'discipline problems' - indistinctively and independent of the level of severity and intensity of the behavior problems (Sørlie, 2000). What is considered as problem behavior may thus vary from teacher to teacher and from school to school. Children exhibiting problem

behaviors in school are frequently referred to as ˙behavior problematic˙ (atferdsvanskelige), ˙trouble makers’ (bråkmakere) or ˙worst cases’ (verstinger). Moreover, student problem behavior has traditionally been viewed as a "within-child-deficit" or "incompetence problem" which imply the position that students show problem behavior because of some individual impairment or weakness (congenital or acquired) or because of a lack of social skills (e.g., Charlton & David, 1993).

In the dissertation, the general terms ˙behavior problems’, and ˙problem behavior’ within the school context are used interchangeably, referring to externalizing student behavior that violates the norms and rules of the school and differ from age-appropriate behavior in a way that it; a) inhibits the student’s own or classmates’ academic or psycho-social development, b) impedes positive social transactions between students or between students and teachers, c) hampers the teaching and learning activities, and/or d) injures or harms other students, teachers or properties (Ogden, 1992; Sørlie & Nordahl, 1998). The terms and definition were chosen in accordance with teachers' perceptions of behavioral problems as they appear in school. Emphasis is put on observed student behavior rather than on diagnostic criteria, individual student characteristics, feelings, or intentions.

Theoretically, the definition builds on a combined social learning perspective, contextual perspective and transactional perspective on student problem behavior. Social learning theory contends that most human behavior, including problem behavior, is learned (Bandura, 1977). Individuals learn from each other through observing others’ behavior, attitudes and outcomes of the behaviors (e.g., positive wins such as acceptances, attention, avoidance of demands or negative costs such as punishment and rejection), which later serve as a guide for action (how to behave) (Bandura, 1977).

The contextual perspective (sometimes referred to as a systemic or bioecological perspective) involves that "each individual child is embedded in a number of systems, notably

family and school, and that the individual´s behavior can only meaningfully be viewed in that sort of contexts" (Charlton & David, 1993, p. 11). According to the bioecological perspective (Bronfenbrenner & Morris, 2007), proximal processes (i.e., interactions between individuals in the microsystem such as family, school, and peers) are the primary engines of human social development. Viewed from a transactional perspective, there is a bidirectional interchange between children and their social contexts, in which both children themselves and the environment will develop and change from the ongoing interactions. As put forward by Stattin and Magnusson (1995), the contextual perspective should explicitly be reflected in how we define, assess and explain behavior problems, as well as in our efforts to prevent or treat behavior problems. The transactional processes may positively or negatively reinforce children's behavior and social development, depending on the contribution from the children themselves and the environment (Sameroff & Mackenzie, 2003). In order to foster children's development of prosocial behavior, the proximal processes should be predominately positive, consistent and supportive, and occur on a fairly regular basis over time (Davis, 2003).

In line with the definition, theoretical foundation, and main objective of N-PALS to prevent and reduce the prevalence of problem behaviors in school, the current dissertation focus on behavioral incidents occurring within and outside the classroom perceived as challenging, negative, harmful or unwanted by the school staff, rather than on assessments of individual student behavior and diagnostic criteria.

1.4 Gender Differences in Behavior Problems

Several meta-analyses of gender differences in the *prevalence* rates of behavior problems in children and youth, including Card, Stucky, Sawalani, and Little (2008), conclude that boys typically show more externalizing problems than girls. The gender differences are larger for physical than verbal aggression (overt). Conversely, indirect or covert aggression (e.g.,

excluding from activities, status violation, spreading gossip, relational conflicts) are more typical of girls (c.f., smaller gender differences) (Card et al., 2008). Likewise, internalizing problems (e.g., anxiety, fearfulness, shame, guilt, sadness/depressive mood) show a marked female preponderance, particularly during adolescence (e.g., Card et al., 2008; Rescola et al, 2007). The two forms of aggression (overt, covert) are highly interrelated, more so among boys than girls, and among older than younger children. Indirect aggression is often associated with higher prosocial skills and internalizing problems, while direct forms of aggression are more strongly associated with low social skills, low peer acceptance and externalizing problem behaviors (Card et al., 2008).

From a developmental psychopathological approach and recent prospective longitudinal studies (e.g., Gutman, Joshi, Parsonage, & Schoon, 2018) we also know that there are gender differences in children's *development* of severe behavior problems. Early-onset disorders such as conduct disorder (CD) are more common among boys than girls (2-4 boys to each girl), whereas adolescent-onset disorders such as depression and anxiety seem more marked in girls than boys (Zahn-Waxler, Schirtcliff, & Marceau, 2008). Although the rates of CD are much higher for boys, the gender differences have diminished since the turn of the millennium (Putallaz & Bierman, 2004).

It has been hypothesized that although the magnitude of behavioral gender differences vary with age (e.g., aggression in girls decline more during the school years than among boys) and contextual influence (e.g., differences in adult's expectations and responses to problem behaviors in boys and girls), they may reflect greater social sensitivity and emotion regulation in girls (Zahn-Waxler et al., 2008). The gender differences may also reflect differences in the social lives of boys and girls (e.g., Keenan & Shaw, 1997; Lahey et al., 2006). For example, girls tend to spend more time indoors, have one or two best friends, and interact in smaller groups than boys (Zahn-Waxler et al., 2008). Boys, on average, report higher self-esteem,

self-concept, and success expectations than girls, and they more often play in larger groups, use direct commands and physical dominance (Zahn-Waxler et al., 2008). Together with advanced physical maturation and better language abilities of girls relative to boys, these and other individual and contextual factors probably make girls more compliant and resilient to problems.

1.5 Prevalence of Behavior Problems in Norwegian Schools

As in the other Nordic countries, there is a lack of systematic screening and longitudinal studies on the prevalence and development of problem behavior in Norwegian schools. However, results from national and international cross-sectional surveys conducted during the last two decades indicate more school problem behavior in Norway than in most other European countries.

In 1998, a representative survey among primary and lower secondary schools, teachers reported concerns about the behavior of 9% of the students (Ogden, 1998). On average, 12% of the teachers considered the behavior problems at their school as serious or very serious, while 15% reported the same in a follow-up study conducted two years later (Lindberg & Ogden, 2000). Another survey conducted in 16 schools (3 upper secondary schools) in two large municipalities indicated that about 11% in primary and lower secondary school and about 20% in upper secondary school showed considerable behavior problems. Of these, 1-2% was defined as students with serious behavior problems (Sørlie & Nordahl, 1998). A research report published in 2009 (Nordahl, 2009) indicated that 4.1% could be defined as students with considerable behavior problems. Additionally, 1.6% was diagnosed with ADHD. A more recent study of behavior problems in primary school (grades 1-7), demonstrated that the prevalence of teacher-reported problem behavior within the classroom context was stable from 1998 to 2008, while *daily* occurrence of problem behavior was

significantly reduced (Sørlie & Ogden, 2014). On the other hand, a slight increase in problem behaviors outside the classroom context was documented, including verbal and physical conflicts among students, lack of concern for others, and running in corridors.

Furthermore, results from the PISA-surveys and the annual Norwegian Student Survey, both building on student reports, indicate relatively stable and high numbers of student problem behavior over time (Kjærnsli & Olsen 2013; Wendelborg, Paulsen, Røe, Valenta & Skaalvik 2012). For example in 2000, 39% of the students (15 year olds) reported of noise and turbulence in most or all class-periods, while in 2003 and 2009, respectively 41% and 40% reported the same (Kjærnsli & Olsen 2013). A decreasing trend has, however, as in the smaller Sørlie & Ogden (2014) study, recently been observed. In 2012, the corresponding number was 29% (Kjærnsli & Olsen 2013).

Taken together, estimations based on school studies (student and teacher ratings) support that student problem behavior was a major challenge in many Norwegian schools at the start-up of the evaluation study in focus of this dissertation – and still is. It seems that about 10-15% of the student body in primary and lower secondary school probably shows moderate behavior problems while 2-3% show serious behavior problems (e.g., White Paper no. 30, 2003-2004). The estimated share of students showing serious problem behaviors match fairly well with the most recent population estimates. Skogen and Torvik (2013) estimated the prevalence of clinically significant behavior problems (CD and ODD) to be 3.5% among Norwegian children aged four to 18 years. Even if this is a relatively modest figure in an international perspective, it amounts to approximately 30.000 children. More seriously, only 0.4% of them are receiving evidence-based treatment or other research-supported interventions (Skogen, & Torvik, 2013).

1.6 Risk and Protective Factors to Externalizing Behavior Problems

Multiple factors in various contexts, in addition to individual characteristics, influence the occurrence and development of behavior problems in children and youth. These factors are called determinants or predictors and can be either negative or positive. Depending on their influence, they are referred to as risk or protective factors. A risk factor is an aspect of the child or of his/her environment that is associated with poor outcomes (Garmezey & Rutter, 1983). Jessor, Turbin and Costa (1998) defined a risk factor as a variable or condition associated with lower likelihood of positive or socially desirable outcomes, and higher likelihood of negative or socially undesirable outcomes (e.g., behavior problems). A protective factor has a reverse effect: it increases the likelihood of positive or successful outcomes and reduces the likelihood of maladaptive outcomes given conditions of risk. It should be noted that a protective factor is not the mere opposite of a risk factor, and its effect may vary dependent on the level of risk. For example, in situations of low risk, a typical protective factor is not particularly protective.

A single risk or protective factor most often has no or only a marginal effect, but the rates of problem behavior increase rapidly with additional risk factors. In other words, it is the accumulation of stress, operationalized as number of risk factors present, in addition to the chronicity and severity of risks (cumulative risk), that make a child vulnerable to externalizing problems (e.g., Deater-Decard, Dodge, Bates, & Pettit, 1998). Opposite, the more support and protective factors a child is exposed to, the more resistant he/she will be to developing behavior problems. Bukowski (1991) suggested that the onset of behavior problems is determined by the interaction between risk factors that predispose a child to the development of behavior problems. He also suggested that protective factors can predispose a child to positive outcomes and buffer a child against development of behavior problems. Said differently, not all children and youths will develop behavior problems even if they have been exposed to several risk factors because the presence of multiple protective factors can lessen

the impact of risk factors. For example, parental support or good social skills may diminish the influence of strong risks, such as having criminal peers and parents with weak parenting skills.

Identification of risk and protective factors can be used to identify children considered at risk (high, moderate, low risk dependent on number of factors) for behavior problems so as to initiate preventive interventions (Hawkins, Catalano, & Miller, 1992). In prevention research, risk and protection theory now is the dominant theoretical framework (Jessor, Van Den Bos, Vanderyn, Costa, & Turbin, 1995). This implies that risk and protective factors should be the targets of intervention. The hypothesis is that an intervention is likely to be more effective the more risk and protective factors that can be identified and influenced (respectively, reduced/ eliminated or strengthened/ increased) (Frazer & Galinsky, 1997). Multi-factor interventions (e.g., N-PALS) that simultaneously address multiple risk and protective factors may thus have greater impact and produce better outcomes than more narrow interventions addressing single risk factors (Hawkins, Catalano. Kosterman, Abbott, & Hill, 1999; OJJDP, 1995). It follows, that the focus should be on risk and protective factors that are *modifiable*, while sociological and structural factors such as a family's socioeconomic status, school size, staff turnover and minority concentration are not particularly useful (although significant predictors of social and academic outcomes), because they largely are outside the control of the schools and not amenable to change by the intervention.

1.6.1 Influential child, family and neighborhood factors. The literature on family influences on children's behavior and social development is extensive, linking proximal factors to children's development of behavioral problems, such as a) child individual characteristics (e.g., adverse temperament, low IQ, impulsiveness, antisocial peers, social rejection by peers, lack of bonding to school, lower academic performance, moral orientation)

(e.g., Gottfredson, Gottfredson, Payne, & Gottfredson, 2005; Kristoffersen, Obel, & Smith, 2014; Payne 2008; Pauwels & Svensson, 2015; Wissink, Dekovic´, Stams, Asscher, Rutten, & Zijlstra, 2014; Welsh, Greene, & Jenkins, 1999), b) conflictual parent-child-relationships, parental stress, single parent home, young age mother, harsh/non-supportive parental discipline, antisocial or socially isolated parents, and more distant factors, such as c) high crime rate, urban/rural location and disorganized neighborhood, (e.g., Murray & Farrington, 2010; Ogden & Hagen, 2014).

Examples of protective factors linked to externalizing problems are: child - adequate language development and academic performance, IQ above average, female gender, prosocial attitudes, high hart rate, high self-control; family - intact family structures, close relationship to at least one parent, medium to high SES, family church attendance, intensive and positive parental supervision; neighborhood – nondeprived and nonviolent neighborhood (e.g., Dubow, Husemann, Boxer, & Smith, 2016; Lösel & Farrington, 2014; Vasallo, Edwards & Forest, 2016).

1.6.2 Influential school factors. A growing body of research has shown that a cluster of *socio-cultural* characteristics of the school organization (e.g., school/class climate, student-teacher relations, disciplinary practice, behavioral expectations, social rules, teacher self-efficacy), are important predictors of between-school variation in level of student misconduct (e.g., Astor, Guerra, & Van Acker, 2010; Gottfredson et al., 2005; Payne, 2008; Thapa et al., 2013; Stewart, 2003; Wissink et al., 2014). For example, Astor et al. (2010) found that students are more likely to experience more violence, peer victimization and punitive disciplinary actions in schools without supportive norms, structures and relations. Results from a large study reported by the Centers for Disease Control and Prevention (2010) showed that a positive school climate was significantly associated with lower levels of student

problem behavior. Several other recent studies report that reduced student violence, aggression, and other types of deviant behavior such as bullying and sexual harassment, are consistently associated with a positive learning environment (e.g., Attar-Schwartz, 2009; Goldstein, Young, & Boyd, 2008; Birkett, Espelage, & Koenig, 2009; Wang, 2009; Wang, Selman, Dishion, & Stormshak., 2010), and positive teacher self-efficacy (beliefs that he or she can positively influence student learning, e.g., Guo & Higgins-D'Alessandro, 2011).

There also is evidence that schools with fair discipline practices and effectively enforced rules and norms experience lower levels of student problem behavior, delinquency, victimization (e.g., Gottfredson et al., 2005), and suspension rates (e.g., Gregory, Cornell, Fan, Sheras, Shih, & Huang, 2010; Wang et al., 2010). Likewise, teachers who interact positively and supporting with their students experience less problem behavior, better classroom quality, and more engaged students (e.g., Brown, Jones, LaRusso, & Aber, 2010). In another study, Gregory et al. (2010) found that consistent enforcement of school discipline and availability of caring adults were related to students' perceptions of school safety. Research also suggests that proactive approaches are significantly more effective in preventing school problem behavior and promoting positive interactions in school than reactive (i.e., reacting to a problem after it has arisen) and passive (e.g., ignoring) approaches (e.g., Soukamneuth, 2004).

In line with results from single studies as mentioned above, several recent research reviews and meta-analyses summarize that the prevalence of student problem behavior varies between schools and indicate that indicators of school climate (e.g., Reaves, McMahon, Duffy, & Ruiz, 2018; Steffgen, Recchia, & Viechtbauer, 2013; Thapa et al., 2013; Wang & Degol, 2016) and teachers' classroom management (e.g., Kopershoek et al., 2016; Marzano, Marzano, &. Pickering, 2003; Oliver, Wehby, & Reschly, 2011) have significant and modest effects on the students' behavior.

As concerns protective school factors, research indicate that a positive school or classroom climate, strong bonding to school, positive teacher-student relationships, supportive and non-deviant peer relations and classroom participation are significant determinants (Lösel & Farrington, 2014; Morrison, Robertson, Laurie, & Kelly, 2002;Vasallo et al., 2016). Some research also indicate that positive perceptions of the ethnic and racial climate in school are associated with lower levels of discipline referrals (e.g., Mattison & Aber, 2007; Shirley & Cornell, 2012), while students in large schools often feel less safe and show more problem behavior and less school connectedness than students in smaller schools (e.g., Lleras, 2008; McNeely, Nonnemaker, & Blum., 2002, Stewart, 2003).

1.6.3 How influential are school factors? Welsh et al. (1999) showed in a smaller study of 11 schools that individual student characteristics, such as school effort, positive peer relations, belief in rules, rewards, and demographic characteristics, accounted for 16% of the variance in school problem behavior, while school and community characteristics accounted for an additional 4 to 4.5%. Gottfredson et al. (2005) found in a considerably larger study (254 schools) that although school structural variables (e.g., size) and neighborhood variables matters, the more *modifiable* aspects of the school and classroom context seem more important. The researchers concluded (p. 432) that

> "The school climate factors influence school disorder net of the effects of the exogenous variables. Specifically, better discipline management is related to lower levels of student delinquency and student victimization, and more positive psychosocial climate is related to lower levels of teacher victimization".

Pointing in the same direction, Korpershoek et al. (2016) concluded on the basis of 54 intervention studies published between 2002 and 2013 that teachers' classroom management practice has an average effect of $g = .22$ on the students' behavior. Hedges *g* is an equivalent to Cohens *d* where an effect size of .20 is considered a statistically small effect. A statistically

small effect of a school-wide or community-wide preventive intervention may, however, be of considerable importance, in that it affects a large number of individuals (Wear & Nind, 2011).

1.7 Prevention and Handling of Problem Behavior in School

Traditionally, the schools' handling of behavior problems have not been characterized by preventive or universal approaches, evidence-based practices or early interventions for at-risk students (e.g., NFR, 2014; Official Report 2009:19; Ogden, 2000:2008; Report to the Storting no. 16, 2006 – 2007). For example, the Official Report no. 19 (2009) recommended corroboration of the schools' commitment to early efforts and preventing interventions. The recommendation was justified by that Norwegian schools generally are characterized by: a) a lack of focus on student outcomes, b) lack of knowledge and use of research-supported measures, c) supportive learning and developmental interventions are implemented too late, and d) a lack of strategies to follow students who fall behind. Moreover, schools have been criticized for individualized and pathological perspectives on student problem behavior while excluding contextual and relational perspectives (Ogden, 2000), and also for providing too narrow measures to such multi-determined problems (Sørlie, 2000).

The need for early interventions and the importance of more effective and scientifically supported measures in the interest of the students have during the last years repeatedly been underlined in several other departmental and inter-ministerial guidelines as well as by scholars (e.g., Arnesen & Sørlie, 2012; Nordahl et al. 2006; Official Directive Q-16, 2007; Official Report 2010:7; Official Report 2011:18; Official Report 2015:2; Ogden, 2012; Report to the Storting no. 30, 2003-2004; Report to the Storting no.16, 2006-2007; Report to the Storting no.31, 2007-2008; Report to the Storting no.18, 2010-2011).

Unfortunately, no recent nation-wide reports or representative data exist for an overview of measures provided to students showing behavior problems in Norwegian schools.

In 1998, Ogden's representative survey among 349 randomly drawn schools showed that of the 19 proactive and reactive reactions listed, the most frequent ways to handle student problem behavior were:1) praising positive behavior, 2) rebuking rule breaking behavior, and 3) giving extra support and encouragements. About 80% of the teachers reported using such strategies on a regular basis. The three least frequent reactions were 1) imposing extra work, 2) placing students in another class and 3) expelling students from school (Ogden, 1998). Sørlie and Nordahl (1998) noted, however, in their study that the instruction offered to 'behavior problematic' students generally appeared as poorly organized and non-systematic with a main focus on their academic shortcomings and learning difficulties rather than on their behavior problems. Later studies also have shown that Norwegian schools have difficulties in offering effective measures and including students at risk of severe behavior problems in regular classes and schools (e.g., Jahnsen, Nergaard, & Flaatten, 2006; Markussen, 2002; Markussen, Strømstad, Carlsen, Hausstätter, & Nordahl, 2007; Nordahl & Sunnevåg, 2008; Ogden 2000).

Special needs education is among the most frequently used measures to meet student problem behavior. It has been argued that 40-50% of all special education needs students (SEN students) in Norway have moderate to severe behavior problems (White paper no. 30, 2003–2004). In a study involving 106 schools, more than 88% of the students in 5th to 10th grade with behavior problems (with or without an ADHD diagnosis) received special needs education (Nordahl & Sunnevåg, 2008). Having an assistant in class (often not qualified as teacher or special education teacher) is the most common type of special needs education, followed by individual instruction (in Norwegian or Mathematics) outside of class and small group instruction (Dalen, 2006; Wendelborg et al., 2015). Evaluations of special needs education in Norway indicate that the students' benefits are questionable and random (e.g., Markussen, Strømstad, Carlsen, Hausstätter, & Nordahl, 2007; Nordahl & Sunnevåg, 2008).

Nordahl & Sunnevåg (2008) found that SEN students, in particular those with behavior problems, significantly got or gave the most negative scores of all students on all variables tested in the study (student-student relations, student-teacher relations, work effort, motivation to learn, academic achievement/grades, student-student relations, student-teacher relations, social skills, and school well-being).

1.7.1 Effective school interventions. By the time of the planning and initiation of the current evaluation study, several meta-analyses and systematic reviews of school-based interventions targeting student problem behavior and mental health problems had been published (Adi, Killoran, McMilan, & Stewart-Brown, 2007; Beelman & Lösel, 2006; Catalano, Berglund, Ryan, Lonczak, & Hawkins, 2002; Garrard & Lipsey, 2007; Hanh, 2007; Horowitz & Garber, 2006; Mytton, DiGuiseppi, Gough, Taylor, & Logan., 2002; Scheckner, Rollin, Kaiser-Ulrey, & Wagner, 2002; Shucksmith, Summerbell, Jones, & Whittaker, 2007; Sørlie, 2000; Wilson & Lipsey, 2007; Wilson, Lipsey, & Derzon, 2003). All concluded that school interventions are in general effective. More recently, Weare and Nind's (2011) mega-analysis of 52 systematic reviews and meta-analyses summarized research on effectiveness and characteristics of school-based interventions conducted across the last 20-years. Most reviews (46) focused on universal interventions, that is, targeted all students, including those without problems, while some additionally explored impacts for students at elevated risk of conduct problems.

1.7.2 How effective are school-based interventions? In terms of specific impacts, the average main effect of school-based interventions on externalizing problem behavior (including bullying and violence), positive mental health, mental health problems and disorders, and prosocial behavior are generally small (ES = 0.1 -0.22), but markedly higher

for high-risk students (ES = 0.21–0.35). The average impact on social skills is moderate to strong, small to moderate on school commitment and academic achievements, while moderate on the family and classroom environments (Weare & Nind, 2011). According to Weare and Nind (2011), better outcomes for high-risk students are probably due to a ˙ceiling effect' for students without obvious problems (i.e., less room for improvement than among high-risk students frequently showing problem behaviors). Cognitive–behavioral interventions consistently show larger effects on the students' problem behaviors than average (mean ES = 0.5), while peer-based and group-based interventions in which difficult children work together, generally have negative (adverse) effects (e.g., more bullying, victimization).

Weare and Nind (2011) emphasize that although the effects of school-based prevention measures may be small to moderate in statistical terms (ES = 0.1-0.5), they should be interpreted as important and relatively large in the real-world. For example, as estimated in practical terms by Durlak and Wells (1997), the average student exposed to universal prevention surpassed the performance of 59% - 82% of the control groups, which reflects a 8% - 46% difference favoring the intervention groups (Durlak & Wells, 1997). Durlak Weissberg, Dymnicki, Schellinger & Taylor (2011) calculated that the effect sizes from 207 social, emotional and learning interventions on average contributed to an 11% improvement in achievement tests, 25% improvement in social skills, and a 10% decrease in externalizing problem behavior, anxiety and depression - and that effects held up for at least six months after the intervention. The effects of universal school interventions seem similar to, or sometimes larger than those achieved in treatment programs for serious behavior problems in children and youths, such as Parent Management Training- the Oregon model and Multi Systemic Therapy (e.g., Ogden & Hagen, 2008; Ogden & Halliday-Boykins, 2004).

1.7.3 Intervention characteristics. Weare and Nind (2011) summarized the characteristics

of the most effective school interventions, characteristics that match well with the N-PALS model:

a) balance between universal and targeted approaches,
b) early intervention (i.e., starting early with the youngest and continuing with the older),
c) teaching of expected school behavior and social skills,
d) focus on positive behavior and mental health rather than on behavioral and emotional problems,
e) operating for a longer period of time,
f) coordinated and joint staff practice within a multimodal and school-wide approach with features such as changes to the curriculum (e.g., teaching skills, linking behavior and academic learning, used active rather than didactic teaching methods, employed interactive methods such as games, role-plays and small group work), improvement of the school culture (ethos), teacher training, cooperation with parents, parenting training, community involvement, and coordinated work with outside agencies (e.g., child welfare, educational-psychological service),
g) high implementation quality (i.e., completely and accurately implemented).

In the following, elaborative comments are given on some of the intervention characteristics.

1.7.4 Balancing universal and targeted approaches. According to Mrazek and Haggerty (1994) preventive interventions can be divided into three subtypes: universal (targeting all students or an entire population, irrespective of risk), selected (targeting individuals or subgroups of individuals with elevated risk), and indicated (targeting high-risk individuals). Interventions may also be blended, combing universal and indicated interventions or universal, selected and indicated interventions (e.g., N-PALS). In schools, universal interventions to prevent behavior problems include practices that should be used by all teachers and school staff to support desired behavior for all students in order to prevent them from becoming at elevated risk requiring more targeted, intensive and resource demanding interventions.

Because better effects for high-risk students are consistently found (independent of type of intervention), some have argued that greater emphasis should be placed on targeted interventions (e.g., Horowitz & Garber, 2006), while others argue that universal interventions provide a more efficacious environment for working with students at moderate to high risk (e.g., Diesktra & Gravesteijn, 2008; Diekstra, Sklad, & Gravesteijn, Ben, & de Ritter, 2008). In their review of research, Adi et al. (2007) concluded that the outcomes appear stronger when universal and targeted interventions are offered in combination, but that the question of optimal balance requires more investigation.

1.7.5 A collective school-wide approach. Several reviews (e.g., Adi et al., 2007; Browne, Gafni, Roberts, Byrne, & Majumdar, 2004; Catalano et al., 2002; Wells, Barlow, & Stewart-Brown, 2003) conclude that mobilizing the whole school as an organization is necessary for the effectiveness of preventive school-based interventions. Weare and Nind (2011) recommend that one should move beyond the sole individual, classroom or curriculum focus, which have been the leading approaches, and embed such work within a school-wide, complex, and multicomponent approach involving a wide range of people, agencies, methods and intervention levels. An effective systemic or school-wide approach to prevent behavior problems and school failure put emphasize on the school staffs' agreement on their *collective* values, norms, expectations and understanding, and on co-joint actions and methodological strategies. For example, an English study showed that schools with a mutual policy and joint practices to prevent and handle student problem behavior had a safer and more orderly learning climate than other schools (Sammons, Thomas, & Mortimore, 1997). These schools succeeded in creating a student-centered environment characterized by effective instruction, social concord and positive learning conditions. In his mega-analysis of factors influencing student progress, Hattie (2012, p. 60) also stress the importance of a collective approach

among the teachers in a school:

> …sharing a common understanding of progression is the most critical success factor in any school, without it, individualism, personal opinions, and ˙anything goes' dominate (usually in silence, in staffrooms, but living aloud behind each closed classroom door).

Creating healthy environments by a school-wide approach is in line with the school policy in many European countries and in other parts of the world (e.g., Australia), and has inspired several recent large-scale agency or government-led frameworks, such as Health Promoting Schools (Schools for Health in Europe, 2010), Healthy Schools (Healthy Schools, 2011), Social and Emotional Aspects of Learning (DES, 2010), Good and Healthy School (Paulus, 2009), Mind Matters (Mindmatters, 2009) and Kids Matters (Kidsmatters, 2009). A Norwegian example is Better Learning Climate (2009-2012), a multi-year national area of concentration, initiated and financed by the Norwegian Directorate for Education and Training. According to the evaluation of this school-political initiative, it resulted in, as expected, a more general or joint approach in many schools. The outcomes for the students (school well-being, student- teacher relations, academic challenges, bullying, motivation, academic counseling, and career counseling) were, however, insignificant (Christensen, Midtbø, Helgøy, & Homme, 2013). The results correspond well with results reported from many other agency or government-led innovations (e.g., school reforms) in Europe and Australia. Common to such approaches are that they are deliberately non-prescriptive and promote ˙bottom-up' principles such as empowerment, autonomy, democracy, local adaptability and ownership. Weare and Nind (2011) uphold that the "European style" makes it challenging to achieve ˙hard' outcomes and measurable changes, and that there is a need to balance this style of approach with more focused and prescriptive elements, greater clarity, and a research-supported foundation in line with more demonstrably effective school-wide intervention programs.

1.7.6 Multiple positive and proactive strategies. In addition to a more holistic approach addressing the whole child rather than just seeing him or her as "problem", the most effective school intervention programs are characterized by frequent and systematic use of multiple positive, skills focused, empowering and proactive teaching strategies (Weare & Nind, 2011), such as praise (verbal, non-verbal), encouragement, concrete contingent rewards for positive behaviors (e.g., stickers), clear instructions/messages, positive involvement (i.e., teachers showing care, sensitiveness, openness, friendliness and engagement in their students' lives and well-being) and modelling of and training in expected behaviors. More effective interventions also use active and interactive strategies (e.g., games, role-play, small group work) rather than didactic strategies (e.g., Durlak & Weissberg, 2007).

1.7.7 A team-based approach. The schools ordinary operational organization normally has restricted capacity and competence to carry out comprehensive interventions. However, to succeed, a school's innovative work needs to be led, planned, coordinated, and regularly evaluated in competent and consistent ways. Several meta-analyses and reviews of school-based intervention research conclude that sustained positive changes in the school culture and the students' behavior and achievement *only* should be expected when the majority of school staff and the principal are actively involved in innovative work in a binding way, and that this effort should be governed by a school-internal resource team (e.g., Diextra & Gravesteijn, 2008; Durlak et al., 2011; Wilson & Lipsey, 2007; Weare & Nind, 2011). This cooperative, decision-making and problem-solving team should ideally consist of 5-6 representative and committed members, including the principal, and for example, a parent representative and a representative from the local Educational-Psychological Service. The resource team need to be well informed of and familiar with the content and methods of the intervention, secure access to required resources, and establish a functional system of supports (e.g., Arnesen,

Ogden, & Sørlie, 2006; Arnesen, Meek-Hansen, Ogden, & Sørlie, 2014; Durlak et al., 2011). Several program purveyors offer a) structured training in the program components, some offer b) additional supervision, while few offer c) a clear implementation structure and a system of quality assurance. NUBU as the national purveyor of the N-PALS-model, offers all three elements.

Interventions led by external specialists (e.g., clinically trained psychologists, researchers) can be effective in short-term interventions for high-risk students and in the beginning of a school's process of change, but use of specialists is not sustainable in the longer term and for universal and larger-scale interventions (e.g., Shucksmith et al., 2007). In other words, for new practices to be embedded in the daily life in schools and for positive student impacts to last over time, the school's regular staff must take the leading responsibility (Weare & Nind, 2011). Parents and communities can add to the innovative work in school if they are appropriately involved and if they support and reinforce the intervention efforts and the skills the children are learning.

1.7.8 Rigorous implementation. To ensure that evidence-based interventions work in ordinary practices, the quality of the implementation process is crucial (Sørlie, Ogden, Solholm, & Olseth, 2010). Implementation quality refers to the correspondence between how an intervention is carried out and how it originally was described by the developers and tested through research, without violating the goals, guidelines, and underlying theory (Domitrovich et al., 2008).

Across settings and programs, high implementation quality has repeatedly been found to predict more positive outcomes and is an important moderating variable (e.g., Fixsen, Naoom, Blasé, Friedman, & Wallace, 2005). A review of six meta-analyses which included 542 interventions, concluded that greater effects are obtained when interventions are

implemented as intended (Domitrovich et al., 2008). For example, Durlak and Weissberg (2007) and Durlak et al. (2011) conclude in their meta-analyses that school-based interventions to prevent behavior problems in children and youth are not all effective if they are based on loose guidelines and broad principles, but that high levels of intensity, consistency, clarity, and program fidelity are needed. Likewise, Berkowitz and Bier (2007), who reviewed a wide range of interventions on character education, concluded that there was a clear trend that incomplete or inaccurate implementation resulted in less outcome effectiveness than complete and accurate implementation.

Accordingly, without information on how and to which degree the intervention has been accomplished in accordance with the original version, no valid conclusions of its effects can be drawn (Domitrovich et al., 2008; Fixsen et al., 2005). If new interventions and methods do not bring about the expected results, this may be due to a) insufficient implementation or b) the intervention being less effective than anticipated. Moreover, if implementation quality is not measured or monitored, one does not know where the problem is, nor how to solve it in situations of non-significant results. Dilution or deviation from the intervention may lower the implementation quality, which in turn results in poor outcomes for the target group. Dilution or deviation may be caused by changes in the group of conveyors, curtailment of the treatment, program elements are skipped or added, or because the staff is becoming less careful and accurate during the implementation process.

Although high implementation quality typically gives better results, the harmonization can be so high that it can be said rigid or mechanical. If the guidelines are followed to the letter, it may imply that the intervention is not adapted to local conditions or the participant's situation, and easily lead to resistance and low commitment. Rigid implementation may as weak or partial implementation, lead to irrelevant, inappropriate or ineffective help and support - the opposite of the intent. In other words, *competent* implementation should be the

main aim, in which loyalty to the method is considered against the need to adapt practices to variations in local context and individual needs (Forgatch, Patterson, & DeGarmo, 2005).

To sum up, school-based interventions to prevent and handle student problem behavior are generally found to be relatively effective and of practical importance. Particularly, school-wide interventions with a multi-modal and multi-level approach such as the N-PALS model, are put forward as relevant and promising – on the condition of rigorous and at the same time, flexible implementation.

2. N-PALS - a School-Wide Approach to Prevention of Problem Behavior in School

This chapter gives an introduction to how the SWPBS model originated in the US. Next, the theoretical and philosophical foundation of the model is presented, followed by a description of the model and its core components in the US and as implemented in Norway (N-PALS). A critical appraisal of the model ends the chapter.

2.1 The Origin of the SWPBS Model

The story of SWPBS began in the USA in the early 90´s by the now classic paper, *Toward a technology of 'nonaversive' behavioral support* by Horner et al. (1990). During the 80's a broad-based normalization and inclusion movement had emerged in support of non-aversive behavior management, reflecting the value that people with severe disabilities who exhibit challenging behaviors should be treated with the same respect and dignity, have the same opportunities, and live in the same settings as others. The movement grew as a reaction to the frequent use of segregating, ineffective and dehumanizing strategies which disabled and behaviorally challenging people often were exposed to. Horner et al. (1990) advocated that such behavioral interventions no longer should be accepted, and that there was an urgent need for alternative and more regulated programming. They suggested more effective procedures for managing challenging behavior and that positive and research-validated procedures to teach and promote adaptive behavior should make the basis of the new routine approach. The normalization and inclusion foundation of this approach also embodies that a) students with behavior problems or other disabilities should be incorporated and supported within the ordinary school and learning environment and not be segregated in special education facilities, and b) that the schools systems and practices must be changed to ensure optimal success and inclusion for all students (Carr et al., 2002).

Moreover, escalating youth criminality and problem behavior in American schools during the 1990´s such as fighting, violence, vandalism, truancy, lack of discipline, drug use, and failure to cope with the problems, called for dramatic changes of practice. Walker et al. (1996) advocated the need for more integrated, preventing and effective strategies. They argued that the reactive "incarceration frenzy" characterizing current practice together with the often incomplete and insufficiently comprehensive school approaches, such as punishment, exclusion, containment, and rejection applied after the fact (i.e., reactive strategies), could not stem the tide of youth violence or adequately address behavior problems among school children. Walker et al. (1996) highlighted the preventing but unrealized key role schools can play, and that schools can serve as a lead agency within an interagency approach to more effective handling of at-risk children.

As an alternative to the mostly reactive, segregating, simple and general solutions traditionally applied to complex antisocial and aggressive behavior, and in line with Horner et al. (1990), Walker et al. (1996) described a school-based differential approach characterized by inclusion and differentiated proactive measures. They initially called it the *Effective Behavior Support* approach (EBS, Sugai & Horner, 1994), which later has been refined and referred to as *Positive Behavior Support* (PBS, Carr et al., 2002) or *Positive Behavior Support and Interventions* (PBIS). During the 90's and at the beginning of the new century, several other supplementary papers describing and arguing for the PBS approach were published (e.g., Colvin, Kame'enui, & Sugai, 1993; Colvin, Sugai, Good, & Lee, 1997; Taylor-Greene et al., 1997; Todd, Horner, Sugai, & Sprague, 1999; Sugai & Horner, 1994; Walker et al., 1998; Walker, Zeller, Close, Webber, & Gresham, 1999). In line with Walker et al. (1996, 1998, 1999), Sugai and Horner (2002) underlined the importance of a whole-school approach and now referred to PBS as the *School-Wide Positive Behavior Support* model (SWPBS). This approach should:

> "a) target the entire school site as well as individual students for assessment and intervention, b) match the intensity and nature of interventions with the severity and intractability of student's adjustment problems, and c) emphasize the fostering of prosocial and safe learning environments for all students" (Walker et al., 1996, p. 196).

By the adoption of the Individuals with Disabilities Education Act (IDEA) in 1997 the significant principles of the PBS approach (e.g., positive behavioral interventions and support, functional behavioral assessment) were established as the new norms for special education policy and practice in the USA. According to the Technical Assistance Center on Positive Behavioral Interventions and Supports, and in line with recommendations from the U.S. Department of Education's Office of Special Education Programs (OSEP), more than 26,000 schools now are implementing the SWPBS model in the USA (www.pbis.org).

2.2 Theoretical and Philosophical Sources

The SWPBS/N-PALS model is based on a broad multi-theoretical and philosophical foundation, drawing inter alia from social-ecological theory (Bronfenbrenner, 1979), developmental psychology (e.g., Sroufe & Rutter, 1984; Masten, 2006), applied behavior analysis (Baer, Wolf, & Risely, 1968), social interaction learning theory and coercion theory (e.g., Patterson, 1982), social learning theory (e.g., Bandura, 1977), cognitive learning theory (e.g., Dodge, 1980), Vygotskij's (1978) socio-cultural vision of learning, social attachment and social control theory (e.g., Hirschi, 1998; Sampson & Laub, 2005). The theoretical foundation of the N-PALS model has been thoroughly described elsewhere (Arnesen, Meek-Hansen, Ogden, & Sørlie, 2014; Arnesen, Ogden, & Sørlie, 2006; Ogden, Sørlie, Arnesen, & Meek-Hansen, 2011). This section thus briefly concentrates on two of the most influential sources from which the SWPBS approach evolved; applied behavior analysis and person-centered values (Carr et al., 2002).

2.2.1 Applied behavior analysis. Applied behavior analysis (ABA, rooted in operant psychology and formerly named behavior modification (Baer, Wolf, & Risely, 1968) served as a critical springboard for the elaboration and development of SWPBS (e.g., Sugai & Horner, 2002; Carr et al., 1999). ABA is a scientific discipline concerned with systematically applying interventions in real-world settings (e.g., schools) with the aim of improving socially important issues, such as behavior problems and learning. ABA is based upon the principles of learning theory, such as contingent use of reinforcement to increase positive (socially significant) behaviors, generalize learned behaviors, reduce undesirable behaviors, evaluate the efficacy of interventions, and make modifications to insure consistent and ongoing progress. Common to all ABA-based approaches are: 1) objective behavior measurement, 2) use of procedures based on scientifically established principles of behavior, and 3) control of the environment (intervention) to allow for objective evaluation of outcomes (Hagopian, Hardesty, & Gregory, 2017).

The early work of Skinner Pavlov, Watson and Thorndike was critical to the development of ABA (Sugai & Horner, 2002). Especially, Skinner's (1938) research on behavior reinforcement gave rise to the field of behaviorism and has been an important contribution to the development of behavior management (Slavin, 2003). Skinner urged that arranging appropriate reinforcement contingencies in the environment is the most critical factor in controlling behavior. A huge body of later research (e.g., Brown, Michaels, Oliva, & Woolf, 2008) support that environmental modifications produce socially significant improvements in human behavior. SWPBS mirrors the use of ABA-based direct intervention practices such as positive reinforcement, stimulus control, antecedent manipulations and contingency management (Dunlap, 2006).

Two other components stemming from ABA which inspired the design of SWPBS are functional behavioral assessments (FBAs) and behavior intervention plans (BIPs; Sugai &

Horner, 2002; Sugai, Lewis-Palmer, & Hagen-Burke, 2000).

Functional Behavior Assessment (FBA) determines the motivation (purpose) of socially significant behavior and the relationship between the behavior and environmental features, and thereby facilitating the planning and design of interventions to change student problem behavior in a desirable direction (e.g., Carr et al., 2002; Schloss & Smith, 1994). In SWPBS, FBA is used as a problem-solving assessment processes conducted by the school's implementation and support team in order to collect relevant information of the problem behaviors and the contexts in which the behaviors occurred (e.g., setting, antecedent, consequences). Based on this information, a hypothesis is developed to summarize and highlight factors related to the problem behavior.

Behavior Intervention Plans (BIPs) are developed based on FBAs for students who do not respond sufficiently to universal or selected interventions (Sugai & Horner, 2002). The purpose of a BIP is to make problem behavior an undesirable choice so that more desirable behaviors can be learned and encouraged in effective ways (e.g., positive reinforcement, behavior modelling) (Sugai & Horner, 2002). Horner and his colleges advocated that BIPs should result in durable, generalized changes in the way the student behaves (Horner et al., 1990). Once in place, the effectiveness of BIPs and changes to the environment (e.g., establishment of systematic positive behavior reinforcement and/or social skills training individually tailored to the child's needs) should be monitored, evaluated, and revised based on data collected after the intervention plan is implemented (Scott & Eber, 2003; Sugai et al., 2000). This function-based approach (FBAs and BIPs) is then used to select and organize empirically supported practices at the school-wide level, the classroom level, in specific non-classroom settings, and/or at the individual student level (see next session).

2.2.2 Person-centered values. As elucidated, operationally defined values and goals

together with empirically validated pedagogical strategies and measures are two of the corner stones in N-PALS/SWPBS. However, the underlying philosophy also "…embraces the idea that while humanistic values should not replace empiricism, these values should inform empiricism" (Carr et al., 2002, p. 6). Accordingly, the decisions of which strategies and measures to use for students at relative risk of developing severe behavior problems should not be solely considered with respect of effectiveness but also with respect to their ability to enhance fundamental values, such as personal dignity and opportunities for choice. From this follows, that person-centered planning is an additional guiding principle of the SWPBS model.

Person-centered planning at the selected and indicated level of supports implies that the development and choices of research-supported measures should be driven by and carefully tailored to the specific needs and risk level of the individual student *and* be considered within the context of inclusion and normalization. That is, individual support plans should be based on needs assessments and FBAs rather than being service-driven. The guiding hypothesis is that if individual needs are met and relevant risk factors are eliminated, or at least reduced, then the student's functioning and quality of life in the school will improve, and problem behavior will be reduced (Carr et al., 2002).

2.3 The School-Wide Positive Behavior Support Model

To be able to effectively prevent and manage student problem behavior, Walker et al. (1996) suggested a multi-level school intervention approach of which the three main purposes are to: 1) identify and reduce known individual and contextual risk factors in school, family, peer-group and community that directly or indirectly trigger or maintain antisocial behavior (e.g., lack of social skills such as self-regulation/self-control, empathy, self-assertion, ineffective parenting and school discipline, conflictual relations to or rejection by peers and adults, low

collective efficacy in school and neighborhood (e.g. Patterson , Reid, & Dishion, 1992; Sørlie & Torsheim, 2010), 2) simultaneously develop protective factors that can contribute to adaptability and successful management of challenges and expectations (resilience, plasticity, skills, perceived efficacy), and 3) provide a "respite" for both at-risk students and the regular school settings.

In addition to the multi-theoretical foundation which emphasize the dynamic influences of the environment on behavior, the SWPBS approach is informed by research on factors that increase risk and factors that buffer or protect against risk for developing problem behaviors or psychological disorders (Flay et al., 2005, c.f., Chapter 1). Together these sources suggest that to adequately affect change and prevent behavior problems in school, it is important to address risk and protective factors at multiple ecological levels (e.g., individual, school, neighborhood) through integrated models of school-based prevention in which multiple, appropriate and evidence-based strategies or interventions are merged into a single school-wide multi-level approach, such as in the SWPBS/N-PALS model

SWPBS is neither a curriculum nor a program but a systemic framework or model that schools can use over time for implementing proactive school-wide discipline (Sugai et al., 2010). The main purpose is to improve the school- and learning-climate and prevent problem behaviors across all school settings. The model has been described as a broad range of systemic and individualized strategies for achieving important social and learning outcomes while preventing problem behavior (Sugai et al., 2015). SWPBS uses pedagogical strategies (e.g., skills training, behavior modelling) to help students develop socially appropriate behavior, while at the same time facilitate changes in broader social systems (e.g., classroom, playground) that influence the students' behavior (Carr et al., 2002).

The core elements and principles of the model are pre-specified and should be successively implemented as described in a handbook. Examples of (and training in) some

relevant evidence-based interventions are also given. However, different from most treatment programs, the list of specific interventions or strategies is guiding and not fixed ˙once and for all'. Schools may use this school-wide framework to guide their selection and implementation of evidence-based behavioral practices to prevent and reduce behavior problems among children and youths (e.g., Walker et al., 1998, Sugai & Horner, 2002) while at the same time being flexibly adapted to each school's needs.

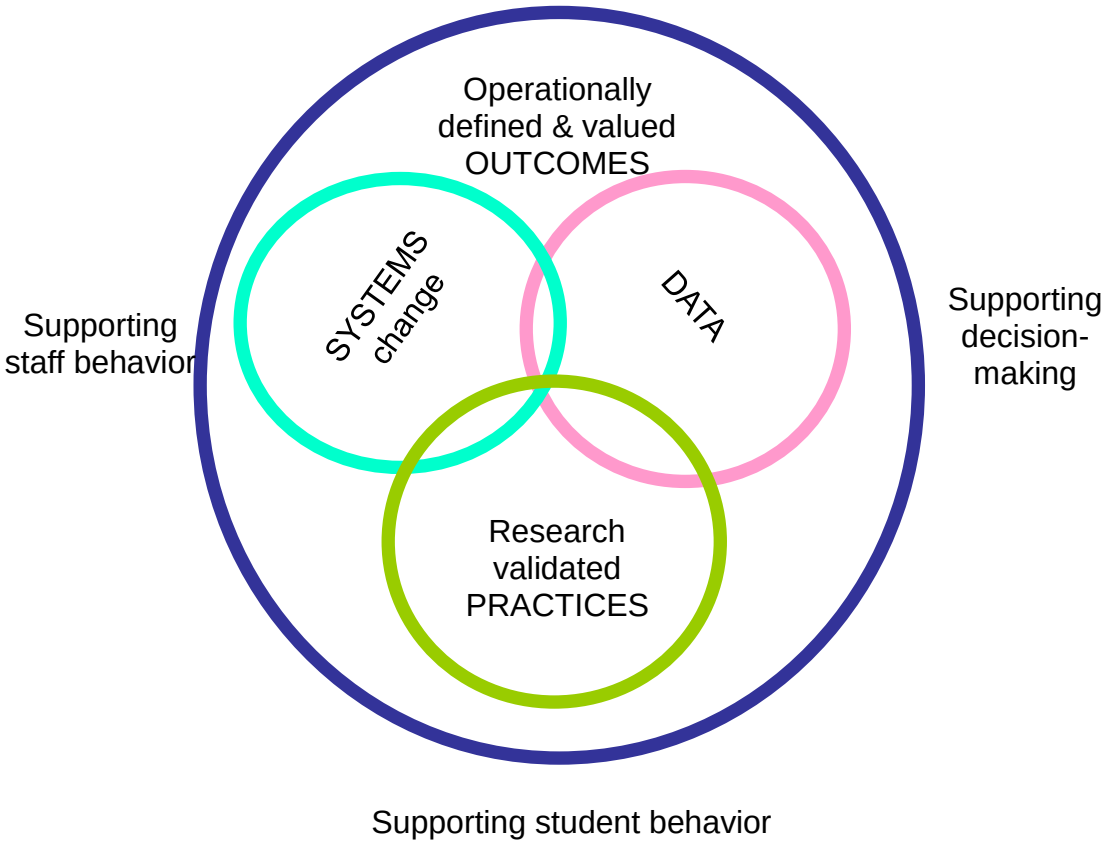
Characteristics of SWPBS include: 1) building a culture within the whole school that will serve as a foundation for both social and academic success, 2) early identification and prevention of problem behavior, 3) directly teaching appropriate social skills to all students, and modifying or rearranging the school context when necessary to prevent problem behavior, 4) using a three-level continuum (universal, selected, indicated) of behavior support practices, and 5) actively using data for decision-making. All students receive the support they need for success based on this continuum of supports and intervention intensity (Sugai & Horner, 2002). Sugai and Horner (2002) also have described SWPBS as characterized by a combination of: a) outcomes that are uniquely defined and valued by the stakeholders like teachers, students, parents and school administration, b) a behavioral and biomedical science of human behavior that provides fundamental principles for the design of support, c) empirically validated practices for achieving identified outcomes in applied contexts, and d) implementation of validated practices in the context of the systems change needed for durable and generalized effects.

Successively over a period of at least three to five years, SWPBS-schools (schools implementing SWPBS) apply the multi-tiered approach using disciplinary data and principles of inter alia applied behavior analysis and social learning theory, to implement predominantly positive and proactive school-wide, targeted, and individualized interventions and supports to improve the school and learning climate, and thereby prevent and reduce the prevalence of

student problem behavior (Sugai et al., 2015).

2.3.1 Critical elements. In order to describe the whole-school PBS model in some more detail, it can be operationalized by integration of four critical elements (Figure 2.1): outcomes, systems, data, and practices (e.g., Sugai & Horner, 2002; Lewis, Barrett, Sugai, & Horner, 2010). To guide the innovative work, each school initially needs to carefully acknowledge and in measurable ways, specify which *outcomes* it wants to focus on (e.g., increased school safety, increased social skills among students, more positive student-teacher relationships, reduced occurrence of minor and major problem behaviors, promoting the staff's competence and efficacy).

Figure 2.1. *Critical elements of SWPBS* (c.f., Sugai & Horner, 2002; Lewis et al., 2010)



First, to be successful and efficient in the development of the school, the operationally defined or specified outcomes must be valued by all significant stakeholders (e.g., staff, students, parents). Second, schools should, as far as possible, prioritize *research-validated practices* and interventions that can maximize the students' and teachers' outcomes. When considering new or different practices and measures, the schools thus should be guided by questions of trustworthiness (validity) of their research foundation, documented effectiveness, and by judgments of relevance.

The strong emphasize on evidence-based practices relates to that teachers and the school system have repeatedly been criticized for ignoring the concept of ˙best practices' and have been shown to resist empirically supported interventions representing more effective and proactive practices (e.g., EU, 2007; OECD, 2007; Ogden, 2008, 2012; Utdanningsforbundet, 2008). Already a decade ago, the glaring disconnection between the availability of proven research-based methods and their effective application by educators was highlighted by several scholars. For example, Kauffman (1996) noted that the educational system is characterized by continuous changes and reforms but little sustained improvement. He argued that the reason for this is that the relationship between reliable and effective practices and their widespread adoption is weak. Malysiak (1997) noted that the first question that teachers seem to pose regarding a potentially effective intervention is; ˙Is it ideologically acceptable or politically correct?' rather than ˙How effectively does it work and what are the limitations?' Instead of relying upon ˙a hodgepodge of activities, unplanned curricula, and conceptually incompatible interventions to accomplish teaching, learning, and management goals', Walker et al. (1998, p. 8, 13) argued that especially within the field of behavioral problems there is a need for

> …a rededication to a profession that a) is research driven; b) focuses upon what works rather than what feels, sounds, or looks good; c) has the potential to enhance the lives of children and youth with emotional and behavioral disorders and their families; and d) contributes to the large public good.

Third, SWPBS relies on *data* rather than on guesswork and unsystematic observations to guide problem solving, decision making and progress monitoring (Sugai, Sprague, Horner, & Walker, 2000). For example, most SWPBS schools are using software programs that can quickly provide visual graphs of the current status and of the development of student problem behaviors over time (e.g., of the occurrence of minor/major incidents, location, time) and of the implementation quality at the respective schools. Several additional survey tools are available online, for example to assess school safety (risk and protective factors in school). For an overview and tools examples, see www.pbis.org

The fourth element deals with *system* change. The schools need to have in place the systems needed to ensure valued outcomes, research-validated practices and data-based decisions. For example, valuation of the effectiveness and relevance of existing organizational working structures and the need for establishing new structures, such as a representative planning and implementation team, individual support teams or specific committees, should be considered, established or refined by need. Likewise, policies and guiding principles (e.g., mission statement, purpose), operating routines (e.g., meetings, communication, problem solving strategies, action planning), resource support (e.g., families, special education, counseling, economic resources, technical support), staff development structures and opportunities (e.g., model-based training and supervision), and administrative leadership (e.g., active participation, visibility, decision making) should at least in the initial phase of the implementation process, be critically discussed and evaluated by the school staff. As earlier mentioned, the SWPBS model allows local adaptations. This also implies that school administrators and teachers should shape or restructure their support systems and goals to address the specific needs at their school. Sugai et al. (2015, p. 17) summarize the four critical elements as follows:

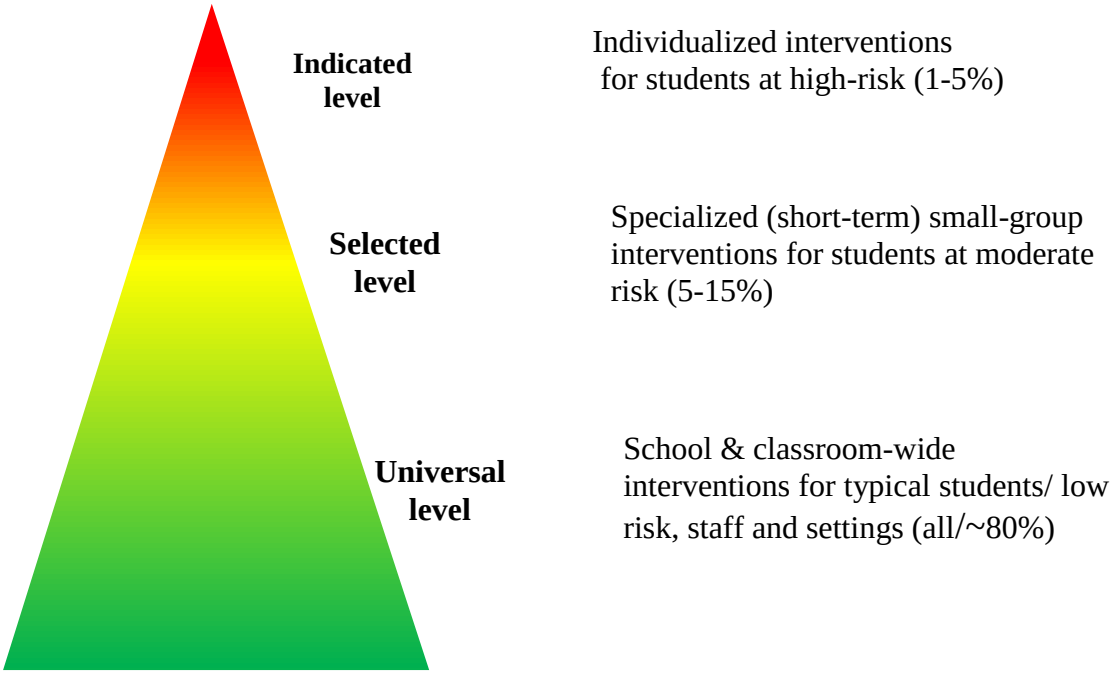
	Description
Systems	Supports that are needed to enable the accurate and durable implementation of practices, efficient use of data, and achievement of outcomes.
Data	Information that is used to select, monitor, and evaluate outcomes, practices, and systems.
Practices	Interventions and strategies that are evidence-based in achieving indicated outcomes.
Outcomes	Behavior and/or academic targets or indicators that are specified, endorsed, emphasized, and monitored because of their social and education significance.

2.3.2 A continuum of support and a multi-level intervention approach. The SWPBS approach builds on the assumption that in every school the student body can be divided into three groups dependent on their needs and level of risk for developing conduct disorders: 1) typical students at no or low risk for problems, 2) students with elevated risk of developing antisocial behavior problems, and 3) students showing signs of escalating or persistent antisocial behavior and involvement in delinquent acts (e.g., Moffitt, 1993; Walker, 1994). Student groups along this continuum of risk are candidates for differing levels or types of interventions corresponding to greater specificity, complexity, intensity, comprehensiveness and expense (Reid, 1993).

To better match interventions to the students' level of risk and needs, one thus should, as far as possible, offer a *continuum* of differentiated and research-validated measures (Figure 2.2), often referred to as primary, secondary, and tertiary interventions (Walker et al, 1996:1998). Within SWPBS, *universal* or primary prevention strategies seek to inhibit student problem behavior to occur or inhibit new cases of conduct disorder by: a) a systematic and collective emphasize on teaching and encouragement of desired student behaviors, b) minimization or preferably removal of school-related factors that promote or sustain problem behaviors, and c) by efforts to maximize school well-being and success for all students.

Selected or secondary prevention strategies seek to lower the rate of students showing moderate behavior problems by reducing or removing risk factors that students bring with them to school (e.g., hot temper, insufficient parenting) by 'bolstering the availability of protective factors' (e.g., intensified positive behavior support, evidence-based programs) (Sugai & Horner, 2002, p. 37). *Indicated* or tertiary prevention strategies applies to individuals who, on objective examination (assessment, screening), are found to be at high risk for severe and escalating behavior problems. Indicated preventive interventions seek to reduce the severity, complexity and intensity of problem behaviors in the individual student's behavioral repertoire or to decrease the amount of disability associated with existing conduct disorder (Mrazek, & Haggerty, 1994; Sugai & Horner, 2002).

Figure 2.2. *A school-wide multi-level prevention approach*



Gordon (1983) pointed to that there is a distinction between indicated interventions and treatment. Whereas the goal of treatment is immediate effect on diagnosable clinical

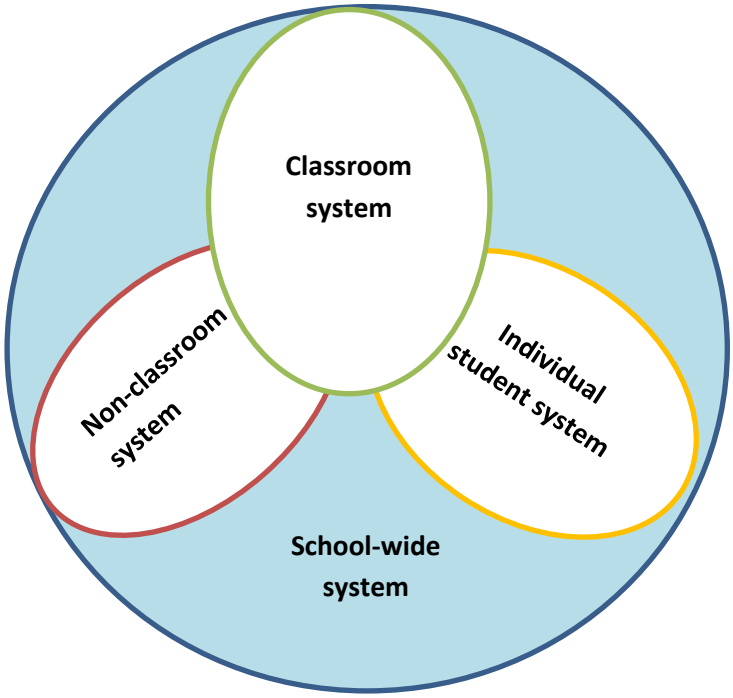
problems, indicated interventions are based on probabilities for later, anticipated symptoms or disability. That is, one cannot be sure that currently observed symptoms on severe problems in fact will develop into clinical conduct, but the odds are seen as high. Accordingly, benefits of indicated prevention are potential and can be delayed for months or even years. The multi-tier or multi-level approach is a hallmark of the SWPBS model that separates it from other school-based interventions.

2.3.3 A multi-system school prevention approach. As urged by Biglan (1995), effective interventions should start with a comprehensive look at the contexts in which problem behavior occurs. In line with this, Sugai and Horner (1994) suggested that the school context should be viewed as a network of four interactive subsystems (Figure 2.3): 1) the school-wide system, 2) the non-classroom system, 3) the classroom system, and 4) the individual system. These subsystems jointly enable students to learn and teachers to teach. Walker et al. (1996) argued that schools should organize their multi-level preventive work in accordance with this four-system perspective to better match the needs, resources, and competence of users. These ideas are built into the SWPBS model.

The *school-wide system* establishes the culture or ecology of the school organization. The focus should be on inclusion of and success for all students by: 1) establishing a statement of policy. Each school's explicit objectives of and rationale for a school-wide discipline structure should be positively phrased, focus on all staff, students and all school settings, and link behavioral and academic outcomes; 2) establishing a set of five or fewer common school rules or clearly defined expectations expressed in few words, for example, Be Safe, Be Respectful, Be Responsible; 3) establishing procedures for teaching prosocial (desired) behaviors that the staff can use to ensure that students know and understand the school-wide rules, expectations, routines, and positive and negative consequences of their

behavior. Basically, the same strategies can be applied as when academic skills and concepts are being taught, for example modelling (tell/show, practice, test), supervision and providing positive or corrective feedback; and 4) organizing and encouraging joint and predictable practices of all school staff members by establishing a common language and understanding of student problem behavior, and common procedures for encouraging expected student behavior, responses to problem behavior/rule-breaking behavior, and for record keeping and decision making (Sugai & Horner, 2002). The procedures should be applied across all school settings.

Figure 2.3. *Multiple systems of school-wide behavior support* (source: Walker et al., 1996).



Correctional procedures should include clear definitions of what is considered minor to major problem behavior in school. To prevent further escalation of aggression and conflict, mild negative responses (e.g., reminds of expected behavior), should dominate rather than intense and high-aversive responses (e.g. angry reprimands, detention, loss of privileges)

(e.g., Walker, Colvin, & Ramsey, 1995). Every school also should establish a common and sustained system for student behavior monitoring and databased decision making, allowing frequent feedback to the school staff about the status of their school-wide discipline efforts (Sugai et al., 2000). A selected group (implementation team) at each school should be able to examine patterns of behavior problems across classes, students, time, locations, behavior types (minor, major), consequences and staff members in order to improve the effectiveness, efficiency and relevance of the intervening efforts (Sugai & Horner, 2002, Sugai et al., 2000).

The *classroom system* constitutes the students' main social and academic learning environment in school. The classroom is first of all characterized by instruction and structured activities, but also by frequent interactions between students and between one or more students and the teacher. Behavior management and routines in classrooms have many parallels to the features of the school-wide system. However, within the classroom context the teachers should ˙maximize their use of behavior management practices' in order to promote a positive and supportive classroom climate (Sugai & Horner, 2002, p. 34). For example, they should have frequent positive contacts with the students. As a rule of thumb, a ratio of at least 4-5 positive contacts or feedbacks for every negative interaction or corrective feedback should be maintained (Arnesen & Meek-Hansen, 2010). For at-risk students (i.e., students showing more frequent and severe behavior problems) the ratio should be 6-8:1 (Latham, 1992). Without a system of proactive and positive behavior support, negative responses to problem behavior are ineffective (e.g., Mayer 1995; Tolan & Guerra, 1994). Teachers also should frequently engage in active supervision so that the students learn that the teachers monitor and evaluate their behavior. Moreover, the teachers should organize the classroom and their instruction in ways that support and facilitate the students' academic learning (e.g., interactive learning, combine "desk-instruction", active skills training or problem-solving, and restricted student teamwork) . To keep up fluency and consistency of their proactive behavior

management practices, teachers need support and opportunities to develop through e.g., regular constructive supervisory feedback, model-based staff training and review of student progress (Sugai & Horner, 2002).

The *non-classroom system* includes the common areas of the school such as playground, hallways, bathrooms, parking lots, and school busses. Such non-instructional areas are typically characterized by large numbers of students, frequent student-student interactions, and limited influence from and presence of adults. Behavior management within the non-classroom system should emphasize supervision that is overt, active, and efficient (e.g., Lewis, Colvin, & Sugai, 2000). The students should be taught how the school-wide rules relate to expected behavior and routines for the specific school areas. Active supervision includes: a) scanning (looking for rule-following and rule-violating behaviors), b) moving (routinely move through locations, especially on areas where there are many students congregate and/or transitions occur), and c) interact (make prosocial and proactive contacts with as many different students as possible, Lewis et al., 2000). Additionally, the staff should provide pre-corrections (structured reminders e.g., on how playground equipment should be used, acceptable hallway behavior, acceptable exit from a bus) that are presented before a student or a group of students enter into situations in which problem behavior previously have been displayed (e.g.; Arnesen & Meek-Hansen, 2010; Arnesen et al., 2006; Sugai & Horner, 2002).

The individual system. Many students who struggle with severe behavior and learning problems do not respond satisfactorily to the general school- and classroom-wide practices. Therefore, they should be offered additional support which is more specialized, intensive, comprehensive, and individually tailored to their needs and level of risk (Sugai, Horner, & Todd, 2000). Sugai and Horner (2002) characterize the individual support system within the SWPBS model as based on: a) an early identification, person-centered and comprehensive

intervention approach, b) functional behavior assessments (FBA) to identify or plan interventions that best fit with the student's needs and risk level (e.g., specialized instruction, curriculum adaptations, research-validated behavioral programs), c) emphasis on social skills and self-management instructions, d) family support and communication, e) multi-disciplinary team-based problem solving and facilitation, f) staff training and implementation evaluation and support, and on g) coordinated cross-disciplinary planning, evaluation and implementation of the tailored interventions.

2.4 The SWPBS Model as Implemented in Norway: N-PALS

In the following, a short description of the SWPBS model as implemented in Norway during the current study period (2007-2012) is given. More detailed descriptions of the model content and implementation structure are given elsewhere, e.g., Arnesen, Ogden, & Sørlie (2006), Arnesen & Meek-Hansen (2010), Ogden, Arnesen, Meek-Hanse & Sørlie (2012), and Arnesen, Meek-Hansen, Ogden, & Sørlie (2014).

As earlier noted, N-PALS is an adapted version of the SWPBS model to Norwegian culture (registered trademark in 2010). The core components of N-PALS as described in a handbook are the same as in the US-model. Except for minor adaptations of the training materials (e.g., pictures, videos, response cards, concepts), no major changes of the original model were made when the SWPBS model was transferred to the Norwegian context. The transportation was done in close cooperation with the University of Oregon.

At the universal level all students were entitled to the following interventions delivered by all school staff members: a) school-wide positive behavior support strategies including teaching of school rules, positive behavioral expectations and social skills, and systematic praise and encouragement of positive behavior (with or with-out reward cards), b) monitoring of student behavior on all arenas of the school by using a school-wide data-based

information system (SWIS), c) collectively applied school-wide corrections with mild and immediate consequences (response cost), d) classroom management skills for teachers, and e) parent information and collaboration strategies.

At the selected level, students who did not profit sufficiently from the universal interventions were identified based on their teachers' perceptions (not on assessments or testing). Neither the original model nor N-PALS includes specific guidelines or cut-off scores for identifying students to selected or indicated supports. The school behavior support team planned inclusive interventions based on what were considered the students' particular needs. Students with moderate behavior difficulties and relatively similar needs, received: a) time-limited small group instruction or training in academic or social topics, or b) the behavioral education program Check-In/Check-Out (CICO) (Todd, Campbell, Meyer, & Horner, 2008), which meet students' needs for positive behavior support and feedback efficiently. The CICO program focuses on improving classroom behaviors through motivation. The student is working on three or four target behaviors which the teacher believes the student needs to work on. The student is daily checking in with a teacher at the start of school to retrieve a goal sheet and encouragement. The teacher provides feedback on the sheet throughout the day. At the end of the school day, the student check out with the teacher, and takes the sheet home to be signed. At check in the next morning, the sheet is returned to the teacher.

At the indicated level, students showing more frequent and severe problem behaviors in school were provided with an individualized and functional behavior support plan based on assessment of the student's school behavior and skills. The tailored interventions could include: a) individual special education (within or partly out of the ordinary classroom) or, b) intensive individual social skills training, e.g. by the use of the cognitive behavioral program Stop-Now-And-Plan (SNAP) (Augimeri, Farrington, Koegl, & Day, 2007). SNAP in school is an evidence-based, manualized, and multi-component group program for minimum five at-

risk children aged six to 11 with serious disruptive behavior concerns (aggression, rule-breaking, conduct problems). The group sessions (45 min.) take place in the classroom setting. SNAP includes a structured curriculum, role plays, and facilitated discussions to help the children learn to identify triggers and make connections between their bodies' physiological responses ("body cues"), thoughts, feelings, and emotional responses. The program helps fostering the development of effective emotion regulation skills to help the children calm down and come up with an effective plan of action. To ensure success, these plans need to: a) keep the children's problems small, b) make them feel like a winner, and c) avoid hurting anyone, anything, or themselves (Augimeri, Walsh, & Slater, 2011). During the study period, parent supports also were available on need as an indicated intervention (i.e., counseling from a therapist based on the parenting principles of the treatment program Parent Management Training (PMTO), Reid, Patterson & Snyder, 2002). However, none of the schools implementing N-PALS asked for this intervention.

2.4.1 National implementation structure. As with the core model elements, the basic training and implementation features are common to the Norwegian and US version of SWPBS, both guided by the PBIS Implementation Blueprint, made available in 2002 (latest version; OSEP, 2015). The only difference is a national system of supports and quality assurance in Norway. Comparable state-wide systems exist in the USA (e.g., Colorado, Florida, Illinois, Maryland, Missouri, North Carolina, Oregon). The national implementation structure in Norway entails that The Norwegian Center for Child Behavioral Development (NUBU) is the only purveyor of the N-PALS model while several institutions offer SWPBS in the US. NUBU's national responsibility for N-PALS, assigned by The Norwegian Directorate for Education and Training, includes high-quality and standardized model training to local coaches for schools interested in implementing N-PALS. A letter of intent signed by

the municipality of the particular school regulates the commitments. NUBU is responsible for the translation, adaptation and revision of the SWPBS model in Norway, and for the handbook and additional materials. Likewise, NUBU is responsible for the development of a standardized training and supervision plan for coaches and resource teams, and for description and revision of the implementation strategy. Moreover, NUBU is responsible for secure storage of the interventions schools' data collected via SWIS (separate server at UiO) and for encouraging sustained high-quality implementation and exchange of experiences with the school-wide model. Once a year, new schools can apply for access to N-PALS.

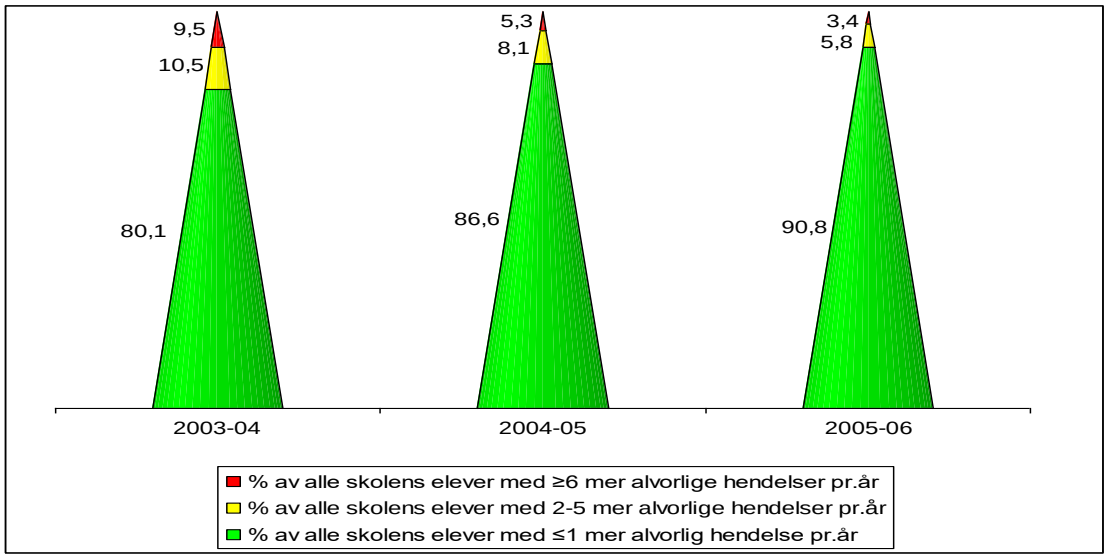
To secure sufficient long term support, each school's readiness for implementation was initially assessed, and approval from at least 80% of the staff was required. Each of the N-PALS schools appointed a resource team (3-4 teachers, the principal, a school psychologist and a parent representative) which was trained on a monthly basis to identify and problem solve the school's problems, effectively collect, analyse, and use data to inform, plan and make decisions on how to best intervene. This school-internal team was also responsible for carrying out, monitoring, and reporting the results at their school. Moreover, the resource team was responsible for offering lectures and training to the rest of the staff in key model features and intervention components (ca. two hours per month per year) and continuously, spend about two hours per week on implementation activities. The teams attended four half-day regional network meetings per school year, in which booster sessions and sustaining implementation activities were provided.

They received local training and supervision from a certified N-PALS local coach for a period of two years (2 hours/10 training sessions per year). The coaches were trained for one year, certified by the national implementation team at NUBU. Most N-PALS coaches were recruited from the School Psychological Services or the Regional Special Education Services. All model training during the study period was free of charge (except for travel costs). Every

intervention school was expected to use the web-based assessment and feedback systems based on nationally standardized assessment tools and routines in order to secure data driven decisions and fidelity.

One of the electronic remedies used, is the School-Wide Information System (SWIS) which on a regular basis allows the staff to enter information on observed student problem behaviors and to generate graphic displays of the aggregated data. For each behavioral incident such as fight between students, verbal "attack" on a teacher or classroom disruption, a standardized form is completed based on the schools own operationalization on what should be considered minor and major behavioral incidents. The form includes the following information: student and staff name, persons involved, location, type of behavior, time point, assumed function of the problem behavior, and consequences (actions taken). Examples on how a school's aggregated SWIS data can be graphically displayed are shown in Figure 2.4-2.6.

Figure 2.4. *Portion of students involved in serious behavior incidents over three school years*



Note: English versions of Figure 2.4 -2.6 are not available.

Figure 2.5. *N reports of minor and major problem behaviors across two school years*

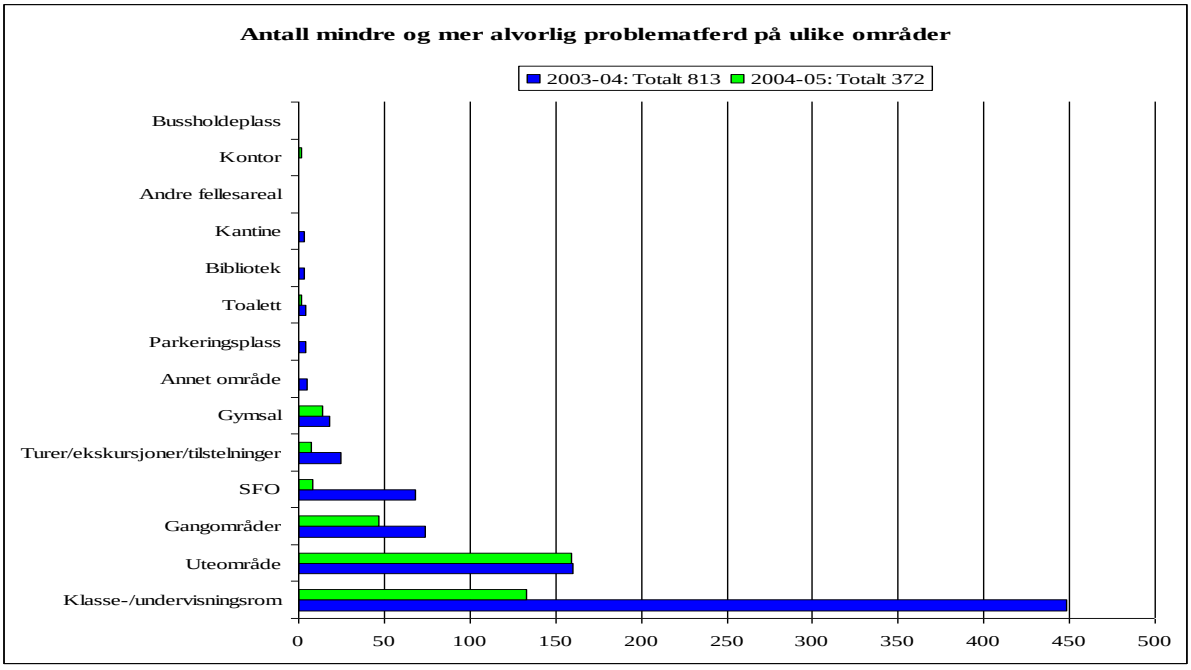
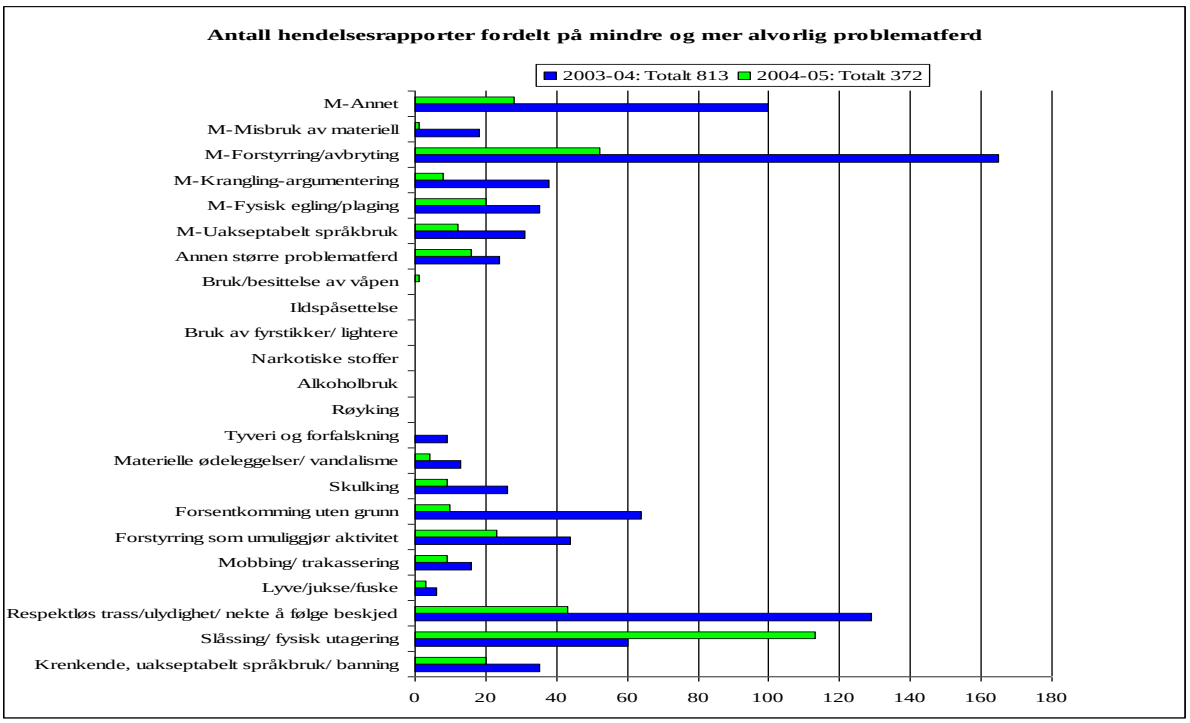


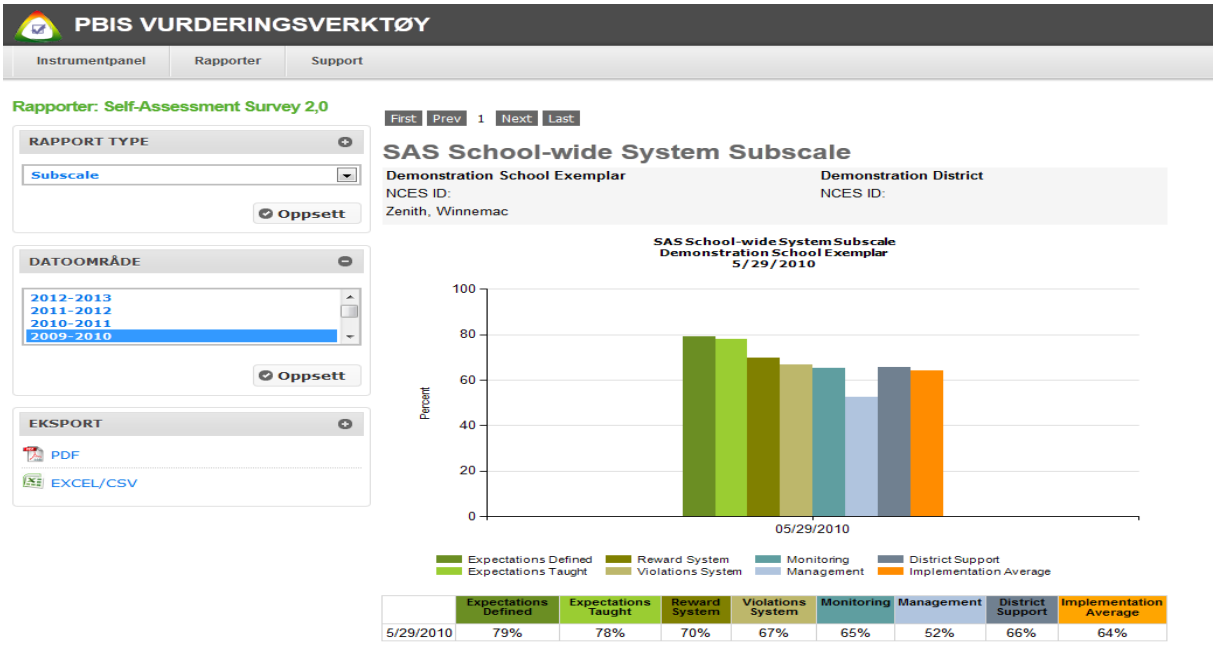
Figure 2.6. *N reports of minor and major problem behaviors on different school areas across two school years*



Another web-based and ʻeasy-to-readʼ tool available for the intervention schools, shows the

degree to which the N-PALS model is implemented at the respective school across years. Implementation data are collected using standardized measures of perceived implementation quality (fidelity), including the Effective Behavior Support Self-assessment Survey (EBS-SAS, Sugai, Horner, & Todd, 2009). This questionnaire is completed annually by all teachers and school staff, and gauges the extent to which the intervention model has been implemented with fidelity at the school level, in the classroom context and in individual cases as well as on common areas like hallways and playground, respectively. An example of how a school's aggregated fidelity data can look like and be downloaded from the web is shown in Figure 2.7.

Figure 2.7. *Example of a school's fidelity scores*



2.4.2 Model developments. During recent years, several refinements and expansions of the SWPBS/N-PALS model have been undertaken to better adapt it to the schools' often parallel behavioral and academic challenges. The behaviorally oriented SWPBS model is grounded in the same principles as the academically oriented '*Response to Intervention*' approach (RTI).

RTI grew from efforts to improve identification practices in special education as a process of systematically documenting the academic performance of students as evidence of the need for additional services after making changes in classroom instruction (e.g., Fuchs, Mock, Morgan, & Young, 2003). According to Fuchs and Deshler (2007, p. 131), RTI in the USA "emerged as the new way to think about both disability identification and early intervention assistance for the 'most vulnerable, academically unresponsive children' in schools and school districts".

Originally being a special education term related to students with learning disabilities, RTI has over the years been integrated with SWPBS and broadened into the general education framework *Multi-Tiered System of Support* (MTSS). Batsche et al. (2005) describe MTSS as the practice of providing high-quality instruction and interventions matched to student needs, monitoring progress frequently to make decisions about changes in instruction or goals, and applying child response data to important educational decisions. According to Kincaid and George (2012, p. 2), the term MTSS is used to describe an "evidence-based model of educating students that uses data-based problem solving to integrate academic and behavioral instruction and interventions" which are delivered across multiple tiers based on the students' needs. In Norway the RTI/MTSS approach has been described under the acronym I- PALS (Arnesen & Meek-Hansen, 2011). However, this extended intervention model focusing on *both* behavioral and academic problems has not yet been implemented due to lack of financial and political support.

Some improvements of the technical support system for high and sustained implementation quality have also recently been carried into effect (e.g., development of an implementation Blue Print, more valid measures of implementation quality such as the Benchmark of Quality). Likewise, changes to the structure of the N-PALS training to make it more efficient, and a revision of the handbook have been carried out during the last years. The N-PALS model has thus far been tested in primary schools only (grades 4-7). However, it

now has been adapted to and piloted in upper secondary schools and day-care centers (Byrkjedal, 2015). Moreover, standardized and validated tests and measurement instruments to identify students with early reading difficulties and social skills deficits recently have been developed (Arnesen, Braeken, Baker, Meek-Hansen, Ogden, & Melby-Lervåg, 2016; Arnesen, Smollkowski, Ogden, & Melby-Lervåg, 2017). Lastly, an additional bullying component has been translated to Norwegian and made available for schools that strive with frequent incidents of bullying (Ross & Horner, 2009).

It should be noted that the model refinements and expansions have happened *after* the initiation of the current study, and none of the participating schools were involved or offered any of the new elements during the study period.

2.5 Critical Appraisals to the SWPBS/N-PALS Model

Recently, OECD (2018, p. 74) recommended that SWPBS should be broadly implemented in Norwegian schools. OECD argues that this approach can;

- particularly benefit boys which represent the overwhelming majority of children referred to special education needs (SEN) programs and the majority of low achievers in the education system,
- complement the current Norwegian SEN model by introducing more low-cost and in-school support to students that do not qualify for SEN support (students in tiers 1 and 2),
- respond to some of the current criticisms of the Norwegian SEN system which is considered too costly and ineffective at responding to the learning needs of students which are below the eligibility margin,
- decrease the number of students that are referred to SEN education by providing them with a lighter school-based alternative.

Previously, researchers have suggested broad-scale implementation of N-PALS in Norway in order to better adapt ordinary education to fit all students, to reduce the level of student problem behavior, to increase academic and behavioral outcomes for all students, and to increase the use of evidence-based practices in our schools (e.g., Ogden, 2014). However, despite its high relevance, sound theoretical and empirical foundation and notable widespread

use (Gage, Whitford, & Katsiyannis, 2018), the SWPBS initiative to system change and effective prevention is neither undebatable nor a "quick fix". When new ideas and practices are introduced in schools, such as the N-PALS model in Norwegian schools, some skepticism and discussion should be expected and also bid welcome.

A question of concern that has been raised, is that the fidelity with which SWPBS has been implemented varies between schools and across time (e.g., Bradshaw, Mitchell, & Leaf, 2010; Borgen, Kirkebøen, Ogden, Raaum, & Sørlie, 2019; Oyen & Wollersheim-Shevey, 2018). In a recent Norwegian register study which included all schools implementing N-PALS by the year 2014 (N = 244), analyses showed that only 18% of the schools had implemented N-PALS with sufficient fidelity (80%) within three years after the initiation of prevention model (c.f., Appendix S4, Borgen et al., 2019). Low, varying or decreasing fidelity over time is a serious threat to attaining and sustaining the significant and positive intervention effects of SWPBS/N-PALS which have been documented in a wide range of effectiveness studies (see Chapter 3).

Insufficient and varying implementation quality (fidelity) is a well-known problem reported in many school intervention studies, and may reflect that the intervention is a) too complex (e.g., multiple tiers, many components), b) demanding to blend into the staff's daily practice (c.f., most Norwegian teachers are not familiar with program-based pedagogical practices, to work systematically, and in coordinated ways over time), or c) that the quality or intensity of the program training, implementation follow-up, or supports over time are too weak in some schools. Other potential limitations of the SWPBS framework and areas of controversy have been suggested. Some of the criticisms made in the USA and Norway are highlighted below.

2.5.1 Critical inputs - USA. Wilson (2015) criticize that the behavioral data collected in the

US intervention schools are exclusively on office discipline referrals (ODRs), and that ODRs rather should be considered reactions to problem behavior than a reliable and valid measure of student behavior. Generally, ODR data do not exist in Norway. Instead, data on specified minor and major problem behavior incidents are collected in the intervention schools via SWIS (Arnesen & Ogden, 2006). Wilson (2015) adds that no data on psychological health, student well-being, or positive behaviors are collected. Similar concerns have been put forward in Norway. In a reply, Putnam and Kostner (2016) suggest that this critique is only partly relevant, in that the model developers (e.g., Horner & Sugai, 2015; Walker, Cheney, Stage, Blum, & Horner, 2005) clearly recommend that multiple sources of data should be used, such as data on suspensions, social skills, academic performance and school climate, rather than solely rely on one source and measures of one sole output variable (e.g., reports of ODRs). Walker et al. (2005) also recommend that standardized screening of all students (completed by teachers) should be a part of the universal tier for early identification of those with behavioral, social, emotional and academic challenges. To meet this criticism, all N-PALS schools are from 2018 expected to carry out universal assessments once a year of a) the students' development in reading through grades 2-5 using the validated measure ORF (Oral Reading Fluency, Arnesen et al. 2016), and b) of the students' social skills in grades 1-6 using the validated measure ESBA (Elementary Social Behavior Assessment, Arnesen et al., 2017).

Moreover, criticism has been raised against one of the theoretical perspectives of SWPBS. Jarmolowicz and Tetreault (2015) argue that applied behavior analysis (ABA) as used in SWPBS does not match the classical APA approach. Putnan and Kostner (2015, p. 99) agree that instead of starting with functional behavior analysis (FBA) at the individual level, SWPBS starts with focusing ˙on the school as the unit of analysis and reserves analysis at the individual level to advanced tiers (specifically, tier 3)'. Loukas (2015) suggests, however, that one may have something to learn from how the ABA principles and methods are included

within the SWPBS framework. She argues that in order to achieve greater societal impact, ABA should be more flexible, multi-component and generic. Critchfield (2015) support this notion and points to that SWPBS probably is the only existing widespread behavioral intervention approach that Skinner would have liked to envision. He argues that the history of ABA is ˙full of perfectly good behavioral interventions that have not achieved wide acceptance in part because they are in conflict with entrenched cultural practices', and adds that when it comes to disseminating interventions at a societal level, may be ABA should become more like SWPBS (Crichfield, 2015, p. 99).

2.5.2 Critical inputs – Norway. In Norway, N-PALS have triggered some debate in the media. Especially, the provision of encouragements and rewards for appropriate behavior (e.g., praise, reward cards, smileys) has been sternly criticized. For example, prof. E. Skaalvik (Dagbladet, 10.07.06) put forward that the problem with the reward system is that it signalizes that you do not expect decent behavior without rewards. Ass. prof. C. W. Beck (Nyskole, 03.01.08) claims that by using rewards, N-PALS makes external motivation the school's basic socialization principle and that the consequence can be ˙…en korrupt befolkning, som gjør alt for penger' (…a corrupt population doing anything for money). Prof. ass. Østrem and Pettersvold (Gjengangeren, 02.01.14) claim that N-PALS contributes to a prohibition culture where "small and large" is defined as rule breaking behavior. They add that in this way N-PALS produces a comprehensive disciplinary problem and conjures up a ghost, which paradoxically can make N-PALS even more popular. Østrem and Pettersvold hold that N-PALS makes all students instrumental because they are being "wrapped up" in positivity and a logic that keeps the teachers on a constant look for rule-breaking behaviors. Østrem (Budstikka, 24.01.14) also states that this way of behavioral teaching is wasted, damaging, and gives associations to dog training. Moreover, in a feature article by a member of the city

council in Oslo, I. M. Thorkildsen (Dagbladet, 29.04. 15) asserts that teachers working with N-PALS are punishing students when they forget their lunch packet, homework, or physical education gear. Additionally, she writes that some kids destroy the class' earned collection of reward cards so that the whole class misses their Friday reward.

In order to get a more valid and broader insight to potentially negative impacts of and experiences with N-PALS, a user survey was conducted among all school leaders with one to nine years of experience with the N-PALS model. The questions were based on the critical inputs put forward in the media, by some professionals, and by some school-political authorities (see e.g., Arnesen, 2011; Arnesen, Meek-Hansen, & Ogden, 2011; NIFU, 2010; Utdanningsdirektoratet, 2010a, b). The user survey indicated strong support to N-PALS in the schools with positive ratings of the schools' benefits and experiences from about 9 out of 10 of the respondents, and almost no indications of negative ripple-effects or supports to the critique that had been raised (Sørlie, Ogden, Arnesen, Olseth, & Meek-Hansen, 2013). Other user surveys involving schools implementing the N-PALS model, point in the same positive direction (e.g., Eriksen, Hegna, Bakken, & Lybn, 2014).

3. A Synopsis of SWPBS Evaluation Research

This chapter gives an overview of the empirical foundation of the SWPBS models' effectiveness *by 2007*. The synopsis is not meant to give a complete overview of all SWPBS-related research, but to sketch out and exemplify the research which inspired the initiation, design, and research questions to be addressed in the current Norwegian study. However, in order to give a more complete picture of the research on SWPBS, results from *meta-analyses* and research reviews conducted during the last years are summarized at the end of the chapter.

A major strength of SWPBS is that it derives from a number of empirically validated principles and practices (e.g., Kern & Manz, 2004). For example, there is substantial literature documenting the principle of reinforcement as an appropriate and effective strategy to promote student behavior (e.g., Cooper, Heron, & Heward, 1987). Likewise, the effectiveness of proactive discipline (e.g., Fenning, Theodos, Benner, & Bohanonon-Edmonson, 2004), pre-correction and active supervision (e.g., De Pry & Sugai, 2002), and functional behavior assessment (e.g., Newcomer, & Lewis, 2004) are well documented[1]. By 2007 and at the initiation of the current study, the evidence-base for the SWPBS components was stronger than the evidence of SWPBS as an effective approach for addressing behavior problems in school (OSEP, 2007). Strong support for core components is, however not sufficient. To hold up SWPBS as an effective system approach, it is critical that the collective practices that define school-wide behavior support and the interventions implemented at universal, selected and indicated level are empirically and socially valid.

[1] For more information, see e.g., Arnesen, Meek-Hansen, Ogden, & Sørlie, 2014; Dunlap et al., 2006; Horner, Sugai, & Lewis, 2015; Lewis, Hudson, Richter, & Johnson, 2004.

At the time the current study was planned, the SWPBS model and intervention research still was in its "infancy", and within the field of Norwegian and Nordic school policy and research, intervention effectiveness and ´standards of evidence' were little-noticed themes. With the exception of Olweus' evaluation of his bully prevention program (e.g., Olweus 1991), very few Norwegian school-based intervention studies had been published.

By 2007, the evidence of SWPBS's effectiveness mainly derived from non-experimental single-case studies (e.g., Bohanon et al., 2006; Lane & Menzies, 2003; Lassen Steele, & Sailor, 2006; Lewis, Powers, Kelk, & Newcomer, 2002; Luiselli, Putnam, Handler, & Freiberg, 2005; McCurdy, Mannella, & Eldrige, 2003; Metzer, Biglan, Rusby, & Sprague, 2002; Scott, 2001), of which few significance tested the reported changes. Indeed, some studies using a more robust research design such as quasi-experiments with pre-post-tests, had been published (e.g., Nelson, Martella, & Marchand-Martella, 2002; Ross & Horner, 2007; Sailor, Zuna, Choi, Thomas, McCart, & Roger, 2006; Sprague, Walker, Golly, White, Meyers, & Shannon, 2001; Sugai, Sprague, Horner, & Walker, 2000; Sørlie, & Ogden, 2007; Turnbull et al., 2002). Of these, the lion's share investigated outcomes after one year of implementation (i.e., Tier 1 or universal level, only). Additionally, results from studies focusing on PBS interventions implemented at the indicated level (Tier 3) and targeting students with or at high risk of conduct disorders, were reviewed as highly promising (e.g., Burke, Hagan-Burke, & Sugai, 2003; Dunlap et al., 2006; Hawken, & Horner, 2003; Kern, Childs, Dunlap, & Falk, 1994; Umbreit & Blair, 1996).

In this extraction of research, examples of studies underpinning core model elements are excluded, as are research on specific strategies and measures used in SWPBS (e.g., team-based school development, behavior reinforcement, office discipline referrals). In accordance with the aim of investigating the effectiveness of the SWPBS model as implemented in Norway, some typical examples of early quasi-experiments and single-case studies are

prioritized. By 2007, one meta-analysis of the research on SWPBS also had been published, while no randomized controlled studies (RCTs) had been conducted.

3.1 Examples of Single-Case Studies

In a single-case study, McCurdy et al. (2003) investigated outcomes of SWPBS in an ethnically and racially diverse inner-city elementary school with approximately 500 students. Office discipline referrals (ODR's) served as the dependent variable. ODR-data were collected monthly over two years. Forms of ODRs contained standard information about each separate offence: the student's and staff member's name, grade, type of offence (e.g., classroom disruption, fighting), and disciplinary action taken. Of comparison purposes, ODR-data also were obtained for the school-year prior to the SWPBS implementation. Brief staff satisfaction questionnaires were administered twice during the project period. At the beginning of the second year of implementation, data on procedural integrity was collected, using the System-Wide Evaluation Tool (SET[2], Horner, Todd, Lewis-Palmer, Irvin, Sugai, & Boland, 2004). The total mean SET score was 82%, indicating high compliance.

Compared to the year before SWPBS (year 1), the number of ODRs per student by the end of the second year of implementation (year 3) was reduced with 46%. As concerns the categories of disruption and fighting, there was an observed reduction of respectively 46.6% and 55%. The mean number of ODRs per student for the classroom and schoolyard was respectively 0.27 and 0.29 in year one, while 0.17 and 0.06 in year three. There was a significant difference in the mean number of all referrals, $\chi^2 = (2, n = 30) = 7.05$, $p = .029$, and

[2] SET is a 28-items measure that evaluates how well a school implements SWPBS and considers seven feature areas (defining of expectations, teaching of behavioral expectations, procedures for acknowledging students and staff, correction procedures, monitoring and evaluation, management of program, and district-level support). SET is considered a reliable measure and exceeds standard psychometric criteria for discriminability, internal consistency, and test-retest reliability. Information is gathered through multiple sources including interviews with administrators, teachers, staff members, and students; reviewing permanent procedures such as school policies, training curricula, and meeting minutes; and examining data systems currently in use.

a significant difference in number of ODRs between year one and year three. The portion of variability in the dependent variable was .24, suggesting a moderately strong relationship between SWPBS and change in number of ODRs. For all but one of six questions measuring staff satisfaction, the mean scores were higher at the end of year two than at the end of year one. Whether the increase in satisfaction with SWPBS (indicator of social validity) was significant, was not tested.

In another single-case study, Luiselli et al. (2005) evaluated the changes in a self-selected urban elementary school (K-5) with approximately 550 students which was implementing whole-school positive behavior supports over a two-year period. The school was followed for three years, including the year prior to implementation. Data on ODR's and suspensions per 100 students were summarized and averaged for each month across the three years. ODRs were reduced from 1.3 (pre-intervention) to 0.73 and 0.25 in year 1 and 2 of the implementation period. The data for suspensions were 0.31 (pre-intervention), 0.25 and 0.20, indicating a small decrease. According to the researchers, the average student percentile ranks on standardized reading comprehension and mathematics tests also improved with respectively 18 and 25 percentage points (data not shown). Staff satisfaction data on two items ("The school discipline plan is effective", and "Student learning time is protected from disruption in our class") were used as indicators of the prevention model's social validity. For both items, teachers responded with increased affirmative responses over the two years SWPBS was in effect. No information on the number of participants at each assessment point were given, and no robust statistical methods were used to analyze the data.

3.2 Examples of Quasi-Experimental Studies

In a quasi-experimental study by Sprague et al. (2002), nine intervention schools and six comparison schools (no interventions) were compared on several measures of school

discipline and school safety, using a pre-post design spanning over one school-year. The elementary and middle schools were recruited to the study by local school administrators based on willingness to participate. The intervention schools implemented a school-wide discipline plan (universal level) based on the Effective Behavior Support model (later referred to as PBS or SWPBS). In addition, the intervention schools implemented the Second Step violence prevention curriculum, a research-supported program applied universally in school (i.e., across all grades) for teaching students social skills (Grossman et al., 1997).

At baseline, a survey was administered to 100 administrators, teachers and parents across the two school groups, measuring 16 risk and 17 protective factors shown to increase or buffer against school violence and discipline problems. No meaningful school differences in perceived school safety were observed. ODR-data were collected both for the year before the intervention and for the intervention year. All intervention schools reported a reduction in ODRs as compared to baseline and showed greater improvements relative to the comparison schools. The average change across elementary schools in the intervention and comparison group was -51% vs. -.075%. For middle schools, the average change was -36%, vs .082 %. Additionally, quarterly reports of school vandalism and their costs were collected in both school groups. The reports indicated less vandalism and costs in the intervention group than in the comparison group. For example for the middle schools, the average numbers of incidents was 8 vs. 11.5 while the annual average cost was 237 dollars vs. 1,550 dollars (one event particularly costly).

A 15-items test to assess the students' ability to define key social skills (e.g., empathy) and respond to vignettes of school-related problems was administered to students in the intervention schools (grades 3-8) prior to receiving social skills instruction and again at the end of the school year. Improvement was observed across all grades (average correct by pre-post was 46% -55%). A measure of implementation quality across the school-wide, common,

classroom and individual student systems was administered to the staff in the intervention schools by post-test (Horner et al., 2004). The results indicated moderate to low fidelity across schools. Supplementary focus group interviews in four intervention and four comparison schools (comprising of teachers, administrators and parents) supported the positive indications of intervention effectiveness. Whether the reported group differences and changes were significant, were not tested.

A second quasi-experimental pre-post study investigated the outcomes of an extended version of SWPBS by comparing seven intervention schools for three years with the remaining 28 elementary schools in one US school district (Nelson et al., 2002). Additionally, a cohort of students exhibiting problem behavior (N = 178) and a non-equivalent comparison group who did not exhibit problem behavior (N = 39) was followed. The extended model consisted of five main elements: 1) a school-wide discipline program, 2) one-to-one tutoring of students in reading, 3) conflict resolution, 4) a video-based family management program, and 5) individualized function-based behavior intervention plan - developed on a consensus-building participatory process. Selection of intervention schools was based on the staff's willingness to focus their efforts on implementing SWPBS. Pre-test was conducted one year before the intervention and post-test at the end of the second year of implementation. To assess the overall outcomes of the intervention model, several teacher and student rated measures, in addition to academic test-scores and administrative discipline data from the year before the intervention and by post-test, were used. ANCOVAs and one-way ANOVAs with pre-test means as covariates were run.

The analyses concerning changes in the school climate revealed significant differences between the two school groups in the pretest-posttest gain scores in number of suspensions, emergency removals, and ODRs (administrative data). The intervention schools showed consistent declines, while increases were observed in the comparison schools. No significant

pre-post changes were found in the students' ratings of school safety as measured by a 26-itmes scale developed for the study, consisting of three underlying subscales (Safety, Relationships, Social Competence). This survey was administered to a random sample of 20 students each in Grades 3 to 6.

Analyses of change in academic achievements indicated significant group differences in the pre-post mean percentile gain scores on a standardized norm-referenced test (Woodcook & Johnson, 1989) on six subjects; Reading, Language, Arts, Mathematics, Spelling, Science, and Social Studies. Except for scores in Mathematics, the students in the intervention group showed greater increases than did their counterparts in the comparison group. Better outcomes in the intervention schools were also observed on a criterion-referenced test indicating the percentage of students who met the state of Washington's desired learning targets. Significant differences were found for two of the three subtests; Reading and Mathematics (not for Listening). As concerns the target students, academic achievement (broad reading, dictation, calculation) and teacher-rated social skills as measured by a 52-item behavior checklist (Behavioral Emotional Rating Scale, BERS), improved significantly in the intervention schools from pre to post-test, while remained stable over time in the comparison schools. A measure of consumer satisfaction administered at post-test, indicated that the teachers in the intervention schools generally reported that a) the techniques and strategies were easy to use, b) the project addressed the educational needs of all students, and c) they would recommend the project to other teachers.

A third example of a quasi-experimental SWPBS-study was conducted by Ross and Horner (2007). Four middle schools implementing the SWPBS model participated in the study. On the basis of measures of the schools' implementation quality scores, the schools were divided into two groups. Two schools with scores above 80% on the School-Wide Evaluation Tool (SET, Horner et al., 2004) were defined as the high implementation group

while the two schools with scores below 80% made up the low implementation group. In addition, five teachers in each school (n = 20) were randomly selected and given measures of teacher stress and teacher efficacy. Data were analyzed using one-way analyses of variance (ANOVAs)

Results indicated a small effect ($\eta^2 = .29$) of the level of SWPBS implementation on perceived teacher efficacy, $F(1,18) = 7.34$, $p < .05$, with higher scores in the high than low-implementation group. Teacher efficacy refers to the teachers' confidence in their individual ability to encourage student learning and positive behavioral change, and was measured by the Teacher Efficay Scale (Gibson & Dembo, 1984), a 30-item measure. Differences in level of implementation did not seem to significantly affect the teachers perceived level of stress, $F(1,18)=.86$, $p =.36$, as measured by self-reports on a 43-item Index of Teaching Stress (Abidin, Greene, & Konold, 2004).

In a forth study, the effectiveness of SWPBS as implemented in Norway under the acronym N-PALS (positive behavior, supporting learning environments and interactions in school), was evaluated with a quasi-experimental pre-post design, including four elementary schools together with four neighboring schools serving as comparison group implementing other interventions than N-PALS (Sørlie & Ogden, 2007; Ogden, Sørlie, & Hagen, 2007). The intervention schools reported that they strived with behavior problems and needed a change of practice (i.e., self-selected), while the comparison schools were invited to the study because of their neighboring location and same size as the intervention schools. Informants were students (N = 735, 7.2% with Norwegian as second language) in 3rd to 7th grade and teachers and staff in after-school services working at least 50% (N = 82). Pre-test was at the beginning of the N-PALS implementation and post-test was 20 months later, at the end of the second year of implementation (Sørlie & Ogden, 2007; Ogden, Sørlie, & Hagen, 2007).

The prevalence of school problem behavior was assessed with three teacher-rated

measures; “Problem behavior in the school environment last week’’ (15 items) and ‘‘Problem behavior in the classroom last week’’ (20 items), both developed by Grey and Sime (1989), and “Behavior problem students in class this year” (15 items, Ogden, 1998). The students’ social skills were measured with the Social Skills Rating System (SSRS, Gresham & Elliott, 1990). The teacher and student forms of SSRS consisted of 30 and 33 items, respectively. To assess the quality of the social-learning context, the Classroom Climate Scale was used (student form = 22 items, teacher form = 14 items, Sørlie & Nordahl, 1998). Perceived teacher collective efficacy (the extent to which a faculty believes in its conjoint capability to positively influence student learning and behavior) was assessed at post-test, using a 12-item short version of the Collective Efficacy Scale (CES, Goddard, 2002). The Total Implementation Quality Scale (TIQS) was developed to measure implementation quality in the intervention schools. The scale was based on the Effective Behavior Support Survey (Sugai, Horner & Todd, 2000) and on SET (Horner et al., 2004). TIQS consisted of 55 teacher-rated items and included two sub-scales referring to the integrity of interventions implemented school-wide and within the classroom context.

Univariate (ANCOVA) and multivariate (MANCOVA) analyses of covariance were run to examine changes over time and to investigate intervention effects. Hierarchical regression analyses were run to test if the magnitude of change was related to program integrity and teacher collective efficacy. Baseline comparisons indicated that the problem situation in the participating schools was more serious than in the average Norwegian elementary school, especially in the intervention schools.

When controlling for the initial levels of student problem behavior, significant intervention effects were observed at post-test, favoring the intervention schools both as concerns problem behavior measured on common school areas such as in corridors, restrooms and on the playground ($F[1,81] = 6.79$, $p = .01$), and in the classrooms ($F[1,81]= 4.67$, p

= .04)]. For example, daily occurrence (i.e., observed one or several times per day last week) of ‘‘verbal attacks on other students’’ declined significantly by 18% in the N-PALS-schools and increased by 6% in the comparison schools, while a reduction of 6% was registered for daily occurrence of ‘‘physical attacks on other students’’ in the N-PALS-schools, compared with a 3% increase in the comparison schools. The two intervention schools with most behavior problems at baseline changed the most over time ($F[3,41] = 3.35$, $p = .03$).

At the classroom level, the largest decrease in favor of the N-PALS-schools (i.e. schools implementing N-PALS) was registered for disturbing behavior, which was the most common form of problem behavior in all eight schools. For example, daily occurrence of ‘‘talking when inappropriate’’ was significantly reduced by 25% in the intervention schools, compared with a 9% increase in the comparison schools (Sørlie & Ogden, 2007).

There also was a significantly larger reduction from pre to post in number of students showing externalizing problem behavior in the intervention group than in the comparison group ($F[1,81] = 6.93$, $p = .01$). No corresponding reduction in number of students showing internalizing problems was found. Likewise, no significant group differences were found in the student or teacher ratings of social skills, except for immigrant students in the intervention group who were rated by the teachers at post-test as significantly more socially skilled than their counterparts in the comparison group ($F[1,47] = 24.2$, $p = .001$) (Ogden, Sørlie, & Hagen, 2007). The reduction in teacher-rated problem behavior and increase in social skills ranged from moderate to large ($d = 0.43$ -1.73). According to the teachers and prior research (e.g., Sørlie & Nordahl, 1998), the quality of the learning environment in class showed a general negative trend in all schools (time effect), but the trend was more pronounced in the comparison schools than in the intervention schools ($F[1,81] = 3.69$, $p = .06$). No intervention effect was registered in the students’ assessment of the classroom learning environment.

Hierarchical regression analyses showed that the reductions in teacher-observed behavior problems in intervention schools were significantly related to how well the N-PALS interventions were implemented, especially at classroom level. Moreover, ANCOVAs, with collective efficacy and the pre-scores of teacher-observed problem behavior as covariates, produced significant group differences, both in prevalence of problem behavior in the school environment ($F[1,81] = 4.08$, $p = .05$), and in number of students showing problem behavior in class ($F[1,47] = 6.72$, $p = .02$). These results indicated that better outcomes in the intervention than in the comparison schools were systematically related to higher collective efficacy (Sørlie & Ogden, 2007).

3.2.1 A meta-analytic review of the SWPBS model's effectiveness. In 2007, Stewart, Benner, Martella, and Marchand-Martella published a descriptive and meta-analytic review of the early research literature on three-tier models of reading and behavior. An in-depth review of 17 peer-reviewed articles was conducted on the similarities and differences between and among three-tier models of reading (N = 5), models of behavior (N = 7), and integrated models (combining reading and behavior; N = 5). The focus in the descriptive analyses was on the student population, intervention level and setting. Additionally, a meta-analysis of 11 of the 17 articles was completed (six did not provide adequate data). Most studies involved elementary school children. Of the behavioral SWPBS-models, five focused on tier one (school-wide or universal) and two addressed all three tiers. Both the reading and integrated models typically occurred at tier two only (selected) and were implemented in the general classroom setting for students considered at risk for school failure.

The meta-analysis produced two weighted mean effect estimates (Z_rs) for each of the three types of interventions. Effect estimates of .10 through .29, 30 through .49, and .50 and

above were considered small, moderate, and large, respectively. For the model with a main focus on behavior (SWPBS), the mean effect on behavior was .28 and .18 on reading, indicating modest impacts of the universally implemented interventions on student problem behavior while small effects on reading performance. For the reading-only interventions, the mean effect on reading was .30 while the researchers were unable to determine any potential effects on student behavior. The mean effect size for the integrated model was .31 on behavior and .53 on reading. It appeared that the integrated model produced the largest gains on reading but the outcomes on behavior seemed not much better than that of the behavior-only model. However, given the moderate to strong correlation between low academic achievement and behavior problems observed in several studies (e.g., Heward, 2006), the researchers stressed the need to integrate reading and behavior into a cohesive model more frequently in the future, and that schools should stop expending resources on ineffective programs or services.

Stewart et al. (2007) added that some limitations of the systematic review should be noticed. First, the number of studies included was relatively small and thus made the results somewhat uncertain. Second, few of the included studies in fact addressed a school-wide three-level approach. Third, many of the studies had weak experimental control and the researchers were unable to account for the quality of the studies (i.e., the degree to which they accounted for threats to internal and external validity). Stewart and her colleges demanded for more rigorous research in the future to determine the effects of school-wide three-tier models for addressing behavior and/or reading problems among students. They concluded that at this point (in 2007), all that could be reliably claimed based on the research was that the three-tier school-wide model had promise.

3.3 Summing Up the Research on SWPBS' Effectiveness by 2007

Generally, results across the early studies of the SWPBS model's effectiveness were positive and promising. Research findings by 2007 suggested that SWPBS stood out as a promising approach for improving the overall school climate by reducing more and less severe student problem behaviors both within and outside the classroom context, such as official discipline referrals (ODRs), assaults, verbal and physical attacks, and school suspensions (e.g., Lassen, Steele, & Sailor, 2006; McCurdy, Mannella, & Eldridge, 2003; Sprague, Walker, Golly, White, Meyers, & Shannon, 2001; Sørlie & Ogden, 2007; Tobin & Sugai, 2005).

Some studies also indicated that implementation of SWPBS, at least for some students and in some schools, can be associated with increased academic achievements because of increased time on task (e.g. Lassen et al, 2006; Luiselli et al., 2005), in addition to increased social skills (Sprague et al., 2002; Sørlie & Ogden, 2007), increased school safety (e.g., Metzler et al., 2001), and increased teacher self-efficacy and collective efficacy (Ross & Horner, 2007; Sørlie & Ogden, 2007). Moreover, a few studies reporting measures of school-level implementation quality indicated that the SWPBS model can be implemented in ordinary schools with fidelity, at least over one to two years, and that better outcomes were related to higher implementation scores (e.g., Bohanon et al, 2006; McCurdy et al, 2003; Ross & Horner, 2007; Sprague et al., 2002; Sørlie & Ogden, 2007). Additionally, high user satisfaction across studies indicated that the social validity of this multi-tier school-based prevention approach is high (e.g., Luiselli et al., 2005; McCurdy et al., 2003; Nelson et al., 2002; Taylor-Greene & Kartub, 2000).

Despite the growing number of evaluation studies by the onset of the current study, several weaknesses of the early research made the evidence of SWPBS's effectiveness questionable. Some limitations are described.

3.3.1 Limitations of the early SWPBS research. To guide the determination of intervention effectiveness, the Society for Prevention Research has proposed a set of criteria, known as the Standards of Evidences (Flay et al., 2005). An *efficacious* intervention has been tested in at least two rigorous trials that (1) involved defined samples from defined populations; (2) used psychometrically sound measures and data collection procedures; (3) analyzed their data with rigorous statistical approaches; (4) showed consistent positive effects (without serious iatrogenic effects); and (5) reported at least one significant long-term follow-up. An *effective* intervention meet all standards for efficacious interventions, in addition to have (1) manuals, appropriate training, and technical support available to allow third parties to adopt and implement the intervention; (2) been evaluated under real-world conditions in studies that included sound measurement of the level of implementation and engagement of the target audience (in both the intervention and control conditions); (3) indicated the practical importance of intervention outcome effects; and (4) clearly demonstrated to whom intervention findings can be generalized. An intervention recognized as *ready for broad dissemination* meet all standards for efficacious and effective interventions in addition to providing (1) evidence of the ability to "go to scale"; (2) clear cost information; and (3) monitoring and evaluation tools so that adopting agencies can monitor or evaluate how well the intervention works in their settings (Flay et al., 2005, p. 151).

First, when applying the Standards of Evidence (Flay et al., 2005) to the early SWPBS effectiveness research, questions can, as noted by Stewart et al. (2007), be raised about the rigor of the study designs to make valid conclusions about effects (see Chapter 4). By 2007, all studies on the effectiveness of SWPBS were either simple single-case studies or quasi-experiments. Second, most of the quasi-experiments were small-scale studies with limited statistical power. Third, there were concerns related to the limited reliability and validity of the outcome measures used in several prior studies. For example, office discipline referrals is

the most commonly used and sometimes the only outcome measure in SWPBS-research. However, ODR data do not provide a pure measure of how much problem behavior is occurring in a given school or a true indicator of the individual student's behavior or behavior change (Metzler et al., 2001; Brandt, Chitiyo, & May, 2014; Sugai, Sprague, Horner, & Walker, 2000). Rather than reflecting externalizing student behavior, one might argue that ODRs reflect the teachers' responses to student behavior (Sugai et al., 2000).

Forth, most of the early studies only evaluated one of the three levels of intervention (universal level) while the possible additional impacts of the other two levels (selected and indicated), were not examined. Thus in large, the effectiveness of the combined three-level model was unknown by 2007. Fifth, except for the pilot-study of SWPBS as implemented in Norway (Sørlie & Ogden, 2007; Ogden et al., 2007), no studies conducted outside the USA had been published. Sixth, the statistical approaches used in the studies conducted before the onset of the current study, generally ignored the hierarchical structure or nesting (clustering) of data (see Chapter 4). Seventh, no studies with an acceptable robust design had investigated the effectiveness of SWPBS over a longer time span than one to two years. Finally, very few studies reported data on how well SWPBS was implemented in the participating schools, how implementation quality statistically related to the observed changes or gave any information on what went on in the comparison schools.

The methodological limitations of the evidence base by 2007 called for more comprehensive, rigorous, and long-term evaluation studies before firm conclusions of SWPBSs effectiveness could be drawn. The promising research findings together with uncertainty about the causal relationship between SWPBS and its outcomes inspired the initiation of a second and more rigorous Norwegian effectiveness study. The following research questions are investigated in Paper 1-5:

Paper 1: *How can we produce valid effect estimates when a preferred randomized controlled design fails?*

- Which are the most significant threats to the internal and external validity in non-randomized experiments?

- According to the methodological literature, what can be done in the design phase to eliminate or at least reduce threats to the validity in the current effectiveness study?

- Did our efforts to strengthen a non-randomized research design result in initially comparable intervention and control groups?

Paper 2: *Three-year impacts of N-PALS on student problem behavior and classroom climate*

- Do schools that implement the three-level N-PALS model demonstrate lower prevalence of staff rated student problem behavior inside and outside the classroom and a more positive learning climate in class (as rated by staff and students) than comparison schools over time?

- Are the intervention outcomes moderated by implementation quality?

Paper 3: *Three-year impacts of N-PALS on teacher efficacy and behavior management practices*

- Does the staff in schools that implement the N-PALS model over a three-year period demonstrate higher perceived efficacy (self-efficacy and collective efficacy) across time than staff in control schools doing ˙practice-as-usual'?

- Does the staff in N-PALS and control schools differ in behavior management (i.e., use of positive behavior support and behavioral correction) as observed by students and staff?

- Are intervention outcomes moderated by implementation quality?

Paper 4: *Immediate impacts of PPBS on student problem behavior, classroom climate, teacher efficacy and behavior management practices*

- Do schools that implement an abbreviated version of N-PALS, the PPBS intervention, demonstrate lower prevalence of student problem behavior over time and better classroom climate (as rated by staff and students) than control schools engaged in ˙practice as usual'?

- Does the staff in schools implementing the PPBS intervention demonstrate more frequent positive behavior supports across time (as rated by staff and students), higher perceived self-efficacy, and higher collective efficacy than employees in the comparison schools

- Are the intervention outcomes moderated by school size, proportion of unqualified staff, program training dosage, and implementation quality?

Paper 5*: Addressing an underlying assumption in SWPBS/N-PALS*

- Do schools with higher perceived teacher collective efficacy scores have lower prevalence rates of student problem behavior observed both in classrooms and at common school arenas than do schools with lower scores on collective efficacy?

- Does perceived teacher collective efficacy predict prevalence of problem behavior in schools over time when significant characteristics of the school, teacher and student body are controlled for?

- Does problem behavior in schools over time predict change in collective efficacy, when significant characteristics of the school, teacher and student body are controlled for?

3.4 Summary of the SWPBS Effectiveness Research Published After 2007

In 2010, Horner, Sugai and Anderson examined the evidence base for SWPBS by conducting a narrative review based on five consideration criteria of 46 peer-reviewed articles published between 2002 and 2009, including two newly published randomized trials (Bradshaw, Koth, Bevans, Ialongo, & Leaf, 2008; Bradshaw, Koth, Thornton, & Leaf, 2009; Bradshaw, Reinke, Brown, Bevans, & Leaf, 2008; Horner et al., 2009). The criteria were: 1) the practice and participants are defined with operational precision, 2) employs valid and reliable measures, 3) grounded in rigorous designs, 4) documents experimental effects without iatrogenic effects, and 5) documents effects. For each of the three levels or tiers of SWPBS the main empirical messages were summarized and exemplified. Twenty of the included articles focused on Tier 1 (universal level, primary prevention), 13 focused on Tier 2 (selected level, secondary prevention), and 13 focused on Tier 3 (indicated level, tertiary prevention).

Horner et al. (2010) stressed that the multiple tiers of integrated practices in SWPBS make the model both complex to define and complicated to evaluate. According to the first criteria, they summarized that the practice and participants were defined with operational precision, in that each of the included studies defined the components of practice with sufficient precision and also provided replicable description of the participating adults, students, or both. Second, they summarized that the research employed valid and reliable

measures, for example by the use of ODRs as an index of school-wide social culture and measures like SET to evaluate implementation fidelity. Third, they argued that a sufficient number of randomized control-group designs now existed and together with several earlier single-case studies, this allowed for assessment of initial effects related to SWPBS practices. Forth, the researchers stated that the existing research documented experimental effects and that no negative effects were found. Further, they argued that together results from the two randomized studies, several quasi-experiments (e.g., Safran & Oswald, 2003; Sprague et al, 2002), and systematic evaluation designs (e.g., Lohrman-O´Rouke et al., 2000, Muscott, Mann, & LeBrun, 2008) formed a consistent pattern demonstrating that SWPBS implementation is possible with improved student behavior. Finally, they summarized that the existing research also documented that SWPBS has sustainable effects with continued levels of reduced ORDs, but that the majority of studies examining longitudinal effects had been descriptive in design. Horner et al. (2010) concluded their review of research that the SWPBS model could be classified as evidence-based due to what they considered sufficient experimental support.

In 2012, Solomon, Klein, Hinze, Cressey, and Peller published a meta-analysis of 20 SWPBS single-case studies published in peer-reviewed journals between 1996 and 2008. Solomon et al. (2012) limited their meta-analysis to outcome variables of problem behaviors and ODRs. A regression-based procedure was used to calculate effect sizes (r^2). A single-case r^2 of .35 refers to a small intervention effect, .65 to a moderate effect, and .90 to a large effect (Parker et al., 2005). All studies examined the effects of SWPBS as implemented at the universal level only, of which twelve measured effects of a multicomponent implementation, whereas eight examined a smaller number of components implemented in various combinations. The most common dependent variable in the reviewed studies was the frequency of incidents of problem behavior, based on ODRs (n = 7) or direct observations (n

= 10) as measured in unstructured settings like playground and hallways (n =10) or in classrooms (n = 8). The duration of the intervention was less than a year in most studies (n = 12) and between one and two years in eight of the studies. Most studies included elementary schools (K-5, n = 13) located in urban or rural districts.

Overall, the effect sizes were positive and ranged from small to medium (r^2 = .27 to .60). The mean effect of SWPBS on student problem behavior as measured by ODRs was .33 and .44 as measured with direct behavior observations. Solomon et al. (2012) stated that since there is significant variability from teacher to teacher and from school to school as to when an ODR occur, the reliability of ODRs are questionable and thus should be supplemented with other measures in future research. The single-case studies indicated that the weighted mean effect size of SWPBS was larger in unstructured school settings (r^2 = . 51) than within the classroom context (r^2 = . 27). The findings were largely consistent across school demographics, although larger effects were observed in urban settings.

The lowest effect sizes were for intervention integrity. Only 60% of the studies reported a quantitative measure of intervention integrity, of which four studies reported implementation integrity in such a way that an effect size could be generated, indicating a mean change in teacher behavior from baseline to intervention corresponding to a r^2 of .40. Given the vulnerability to low implementation quality of large-scale system changes in school, Solomon et al. (2012) suggested that researchers and practitioners should devote significantly more resources on measuring implementation quality in the future.

Moreover, the observed changes across intervention length were small, suggesting that the effects of SWPBS remained fairly constant over time. Solomon et al. (2012) pointed, however, to that future research would add value to the evidence base by conducting more studies that covered at least two full school years. They also emphasized that one should not expect SWPBS to reach maximum integrity until three to nine years of implementation and

that current the lack of longitudinal studies beg the question; How sustainable is the effects of SWPBS over time? The meta-analytic results indicated positive changes both in elementary, and particularly in middles schools, following the implementation of SWPBS. According to Solomon et al. (2012), the imbalance of number of studies conducted in these contexts (n = 13 vs. 3) called, however, for more research in middle school settings as for younger and older students in the future. They also called for future validation of individual components of SWPBS – in isolation and combination, such as active supervision, a token economy, or pre-teaching, and to better understand what combination of interventions best matches different levels of student need. Finally, to qualify SWPBS as evidence-based, Solomon et al. (2012) called for more future research on the full three-level model and that it would aid the practitioners to know what frequency of non-responders may be expected when SWPBS is implemented.

In a forth review, Chitiyo, May, and Chitiyo (2012) examined most of the same studies as Horner et al. (2010), but by using a more sophisticated approach, they came to a different conclusion – a conclusion closer to what Solomon et al. (2012) hold out. Although Horner et al. (2010) stated that the five criteria used for considering SWPBS as an evidence-based model primarily applied to the consideration of individual studies, Chitiyo et al. (2012) argued that Horner et al. (2010) instead applied the criteria to the whole research body. By doing so, they "created a false impression about the solidity of the research behind SWPBS because studies that fulfill one of the quality indicators may not be sited to support the body of research when those studies may actually have fundamental methodological flaws" (Chitiyo et al., 2012, p. 3).

In line with Horner et al.´s suggestion, Chitiyo and his colleagues (2012) decided to base their review on the same proposed criteria but now applied to individual studies published in peer-reviewed articles between 1990 and 2011. Their search generated 34 articles

of which 24 were descriptive non-experimental studies mostly based on the AB single-case design. These articles were excluded from the analysis because “descriptive studies are designed to provide descriptive information about SWPBS and not necessarily to determine experimental effects” (Chitiyo et al., 2012, p. 6). They found that only four of the 10 included studies defined the participants and practice with precision, and that three studies met the criterion of rigorous design while four partly met the criterion. Moreover, eight of the 10 studies employed valid and reliable measures but Chitiyo et al. (2012) were critical to studies using only one outcome variable, mostly ODR. Additionally, they found that seven of 10 studies documented experimental outcomes without iatrogenic effects. The effect sizes ranged from .00 to .90. Only two studies demonstrated both high fidelity (at least 80%) and sustained administrative support as well as reporting social validity measures reporting user satisfaction. Overall, Chitiyo et al. (2012) summarized that only two studies, both RCTs, (Bradshaw, Mitchell, & Leaf, 2010; Horner et al., 2009) met all five criteria for the evidence base for SWPBS. On this basis, the researchers concluded that SWPBS “is quite a promising approach which requires more inquiry with enhanced methodological rigor to be established as evidence based” (Chitiyo et al., 2012, p. 20). They emphasized that future research on SWPBS should pay particular attention to issues like rigor of the research designs, implementation integrity, operational definition of participants and practice, and use of more valid and reliable student outcomes measures.

In 2018, Noltemeyer, Palmer, James, and Wiechman published yet another synthesis of the existing research on SWPBS, however, restricting it to research examining effects of the universal school-wide tier only. The aim of the study was to summarize what is known about the impact of SWPBS not only on student problem behavior but on other outcome variables, such as academic achievement. In the initial section of the synthesis, they summarize that SWPBS in many studies has been associated with reductions in problem

behaviors (e.g., Bradshaw et al., 2010; Flannery, Fenning, Kato, & McIntosh, 2014) and bullying (e.g., Waasdrop, Bradshaw, & Leaf, 2012), increased school attendance (e.g., Freeman et al., 2015; Pas & Bradshaw, 2012), improved student and staff perceptions of safety (e.g., Horner et al., 2009; Kelm, McIntosh, & Cooley, 2014), improved perceptions of teacher efficacy (e.g., Kelm & McIntosh, 2012; Reinke, Herman, & Stormont, 2013), enhanced academic outcomes (e.g., Pas, & Bradshaw, 2012; Simonsen et al., 2012) and social competence for students (e.g., Bradshaw, Wassdorp, & Leaf 2012; Cook et al., 2015), in addition to enhanced school climate (e.g., Bradshaw et al., 2008; Caldarella, Shatzer, Gray, Young, & Young, 2011).

Noltemeyer et al. (2018) identified 77 studies of which 27 were excluded because they did not meet the inclusion criteria (n included = 55, because some articles reported more than one study). The eligibility criteria included; documents impact of SWPBS on at least one outcome variable of interest, involves participants in grades preschool through 12, reports published or unpublished pre-post measures or comparisons to outcomes in a school not implementing SWPBS, uses a correctional, experimental, quasi-experimental, causal-comparative, or a descriptive design albeit without egregious methodological flaws, reports in English, and is conducted since 1990. A multi-stage coding system was applied, including six levels of evidence (*1* = evidence from a systematic review of all relevant RCTs, or evidence-based clinical practice guidelines based on a systematic review of RCTs, *6* = evidence from the opinion of authorities or report of expert committees), and a five-level rating of the overall impact (*1* = unanimously significantly positive results across all outcome variables, *5* = unanimously nonsignificant results across all outcome variables).

Fifty-two of the included studies were conducted in the USA and three in Canada, which on average included 72 schools and 14,753 students. The samples most often had a heterogonous composition with about 50% girls and 50% white. About half of the studies did

not specify the participants socio-economic status (n = 29), type of school (n = 31) or location (n = 25). In those who did, between 50% and 100% of the students had low SES, most schools were urban and public. Respectively, one and seven studies investigated effects of SWPBS in preschool and high school, while effects in elementary and middle (junior high) school were reported in 20 and 10 studies, respectively. Seventeen studies had a multi-grade approach.

A descriptive single-case design was the most frequently utilized (n = 30), while 18 articles reported outcomes from RCTs or quasi-experiments (n = 9 of each), and seven from correctional design. On the positive side, 42 manuscripts (76.4%) reported on implementation fidelity data. In 39 of the 55 cases the evidence was rated at level five (i.e., evidence from a single descriptive, qualitative, or causal-comparative study). Totally 15 outcome variables were examined with ODRs (n = 35), suspension (n = 18) and achievement (n = 14) as the top three, followed by school climate (n = 7), teacher attitudes (n = 5), problem behaviors (n = 5), behavioral strengths/competencies (n = 5), and attendance (n = 5). Among the studies utilizing significance testing (43.6%) , the results suggested a positive overall impact of the universal level of SWPBS, in that the findings related to problem behavior were either unanimously (n = 5) or predominantly positive (n = 9) across all variables of interest. However, as concerns academic achievement, almost 50% of the findings were either unanimously or predominantly non-significant while only 27.3% of the cases reported positive significant outcomes.

A systematic review by Mitchell, Hatton, and Lewis (2018) should also shortly be mentioned. Their goal was to summarize research on the impacts of SWPBS training, and identified five relevant studies with a group-experimental design. Although the fidelity of the training was not measured in any study, Mitchell et al. (2018) concluded that schools can be adequately trained to implement SWPBS with fidelity. No effect sizes were reported.

The last summary of the research published on the SWPBS' effectiveness was published in 2019 by Lee and Gage, which was an update of a systematic review published the year before by Gage, Withford, and Katsiyannis (2018). Different from all prior reviews, this one included all available school, staff, and student-level outcomes, including academic, behavioral, and organizational outcomes, and also unpublished studies and dissertations (to reduce publication bias). Gage et al.'s (2019) search and quality evaluation ended with 32 experimental studies with a randomized or a quasi-experimental design. Of these, 12 were dissertations. Totally 8,781 schools across 10 locations in US and Europe (i.e., Norway) were included in the studies. Twenty-two of the studies reported on implementation quality, which was between 70%-80%. The majority of the studies were conducted in elementary and middle schools, and the average years of implementation were 3.4 years. There was significant heterogeneity for the effect sizes, but the mean school-level effect size for the behavior domain was 0.26, 0.11 for the academic domain, and 0.37 for the organizational domain. Gage et al. (2019) conclude that SWPBS has a statistically significant and meaningful effect on behavior, academics, and organizational health. According to Cohen's "rule of thumb" the ESs are relatively small. However, according to ES benchmarks suggested for the educational field (Bloom, Hill, Black, & Lipsey, 2008; Hedges & Hedberg, 2007; Lipsey et al., 2012; Kraft, 2018), the overall results on the behavioral and organizational domain should be considered large while medium on academic achievement.

Taken together, the systematic reviews published during the period 2008 - 2020 show that SWPBS stands out as highly promising for the prevention of externalizing behavior problems among students, and for improving the organizational health, learning climate, and safety in elementary and middle schools. On the other hand, the evidence on SWPBS's effects in preschool, high school, and on academic achievement seems more limited, mixed or inconclusive. As by 2007, there still is a need for research using robust longitudinal designs

and fidelity measures allowing for more certain causal conclusions regarding the impacts of the SWPBS model. Most research studies published both before and after 2007 have focused on SWPBS' universal tier only, although some small scale studies have demonstrated promising outcomes of selected level interventions (e.g., Hawken, MacLeod, & Rawlings, 2007) or indicated level interventions (e.g., Iovannone, Greenbaum, Wang, Kincaid, Dunlap, & Strain., 2009). The current study is one of the first investigating effects of the *combined* three-level SWPBS model (Humphrey, Lendrum, Barlow, Wigelsworth, & Squire, 2013), and the very first study investigating its effectiveness over several years, and as implemented in an European country.

4. Method

In this chapter methodological considerations and descriptions of the study design, participant selection procedure, data collection procedures, instrumentation, statistical power, and the analytical approach are given.

4.1 Procedures

Multi-informant data were collected from participants in 65 primary schools at six time points (T1-T6) across five successive school years, including a double pretest (conducted respectively four months prior to and at the beginning of the interventions and a new school year), two intermediate tests (after one and two years of implementation of N-PALS), a posttest (after year three when the N-PALS model was expected to be fully implemented), and a follow-up test (after year four with N-PALS). Questionnaires were administered to all students in 4th through 7th grade (9–12-year olds), to their class head teachers, and to all other teachers, assistants, and after-school personnel who were in daily contact with a group of students. Additionally, the principals and a 30% random sample of the parents contributed to the study. Selected data from T1-T5 were used in the analyses presented in the papers included in the dissertation. Data from T6 (follow-up), the parents and other data from the study will be analyzed and published in future papers.

Only students with informed and written assent from their parents were allowed to participate in the study by completing a questionnaire with a teacher present. The completion was conducted during 1-2 lessons, dependent on the age of the students. If needed, the teacher read the items and instructions aloud and explained the procedure to the students. In addition to collecting updated lists of students and staff and parental consent for new students and new cohorts of 4th graders, information on the respondents' status (long-term leave of absence, quit

or changed school) were collected each year prior to the data collection. Assent from school personnel was obtained when completing the questionnaires. It should be added that after the 3rd data collection, a few schools reported problems in motivating some of their staff members (in particular, after-school personnel) to complete the questionnaires. To prevent exhaustion and attrition of respondents, we therefore decided to skip the staff-questionnaire at T4 (i.e., after two years with N- PALS); a decision met with gratitude from most schools.

All questionnaires were available in Norwegian both on the internet and on paper. Questionnaires on the internet were distributed by an external contractor with broad experience with annual nation-wide data collections administered by the Norwegian Directorate of Education and Training. To standardize the assessment procedure, written instructions were provided for all participants. The parent consent letter and questionnaire were available in English, Urdu, Somali, and Bosnian, which are four of the most frequently spoken foreign languages in Norway. The texts were translated and quality checked by professional interpreters.

4.1.1 Recruiting schools to the study. Problems in recruiting schools to the current study were expected because of 1) extra work load for all schools following a national school reform in 2006 in order to increase the academic achievement level among Norwegians students (Knowledge Promotion Reform, www.udir.no), and 2) an observed increase in school-based project activities to promote the learning conditions or to reduce student problem behavior (KUF/BFD, 2000). Accordingly, 40-50% oversampling was decided (i.e., inviting more schools than needed according to a priori sample size estimation, see below). On this background, 91 schools located in 17 strategically selected municipalities were initially invited to participate in an evaluation study with randomized allocation either to an experimental group implementing the three-level N-PALS model or to a control group

implementing an abbreviated version of the N-PALS model, developed by Anne Arnesen and Wilhelm Meek-Hansen at the NUBU, and called the *Preventing Problem Behavior in School* intervention (PPBS)[3]. The goal was an experimental group and a control group with 32-33 schools each. The schools were invited by an e-mail and an information pamphlet to the school leaders.

Nineteen schools (21%) volunteered to participate. However, several of these schools informed that they refused to participate if they were allocated to the control condition. This implied that only nine schools in fact accepted randomization. The four main reasons for schools refusing to participate were (rank order); 1) "We do not have the capacity to implement any program now, even if behavior problems are a major challenge. 2) "Due to the preplanning and implementation of the "Knowledge Promotion Reform", the teachers show clear signs of exhaustion and have expressed resistance to yet another innovation", 3) "We are implementing another school project" or "We are already participating in another program evaluation study", and 4) "We have tried PPBS-like programs before, and do not think such interventions can produce sufficient and sustained effects". Accordingly, and as described in Paper 1 and further argued in the next paragraphs, the preferred randomized controlled test design (RCT) had to be abandoned. After considering several alternative designs, a cluster-based quasi-experiment was seen as the most appropriate (NEC, nonequivalent control group design). A deliberate attempt to control the recruitment of schools to this study was aimed at by a stepwise recruitment procedure (see Figure 4.1 for a recruitment flow chart).

Step 1: A stratifying and matching procedure was used prior to inviting schools which implied that all primary schools (grades 1-7, N = 165) in 17 strategically selected municipalities were divided into three strata based on school size (small schools, N < 150

[3] The abbreviated intervention also has been called the "Brief School-Wide Positive Behavior Support" program (BSBS) (Ogden, Sørlie, Arnesen, & Meek-Hansen, 2012; Sørlie & Ogden, 2014; Sørlie, Ogden, & Olseth, 2015).

students, medium schools, N = 151-250 students, large schools, N < 251 students). The municipalities were located in the Southern and Western parts of Norway and were selected because they were the only ones with access to a local N-PALS coach in addition to a certified PMTO therapist, which was determined by the model adapters as necessary pre-requisites for implementing the N-PALS model.

Step 2. Schools that were actively implementing other school- or community-wide programs, such as the Olweus bullying prevention program (Olweus & Limber, 2010), Zero (Midthassel & Ertesvåg, 2008), Respect (Ertesvåg & Vaaland, 2007) or TIBR (formerly EICR, Kjøbli & Sørlie, 2008), were excluded as potential participants to avoid program contamination. Information on the use of other programs was collected by the schools' home-page and validated via phone or e-mail.

Step 3: Additionally, because N-PALS is a demanding model to implement and in order to put the model on the hardest test possible, schools with fewer than 100 students were also excluded as potentials (two exceptions due to lack of alternatives). In total, 39 of the 165 schools (24%) were thus initially excluded, which resulted in an eligible sample of 126 schools.

Step 4: Each school in this eligible sample was assigned a number on a list and then carefully informed and *randomly invited* by blind drawing of numbers from a box as intervention or control schools. Due to the earlier mentioned expected problems in recruiting a necessary number of schools, an approximate 50% oversampling rate was applied. Hence, 52 schools in the eligible pool were invited as N-PALS implementation schools according to a matching 10% - 25% - 65% distribution of small, medium and large schools, respectively (corresponding to the national distribution). Of these, 28 (54%) schools agreed to participate

Step 5: Thereafter, schools in the control pool were matched to the N-PALS group based first on school size and second on geographical location (distance to an intervention

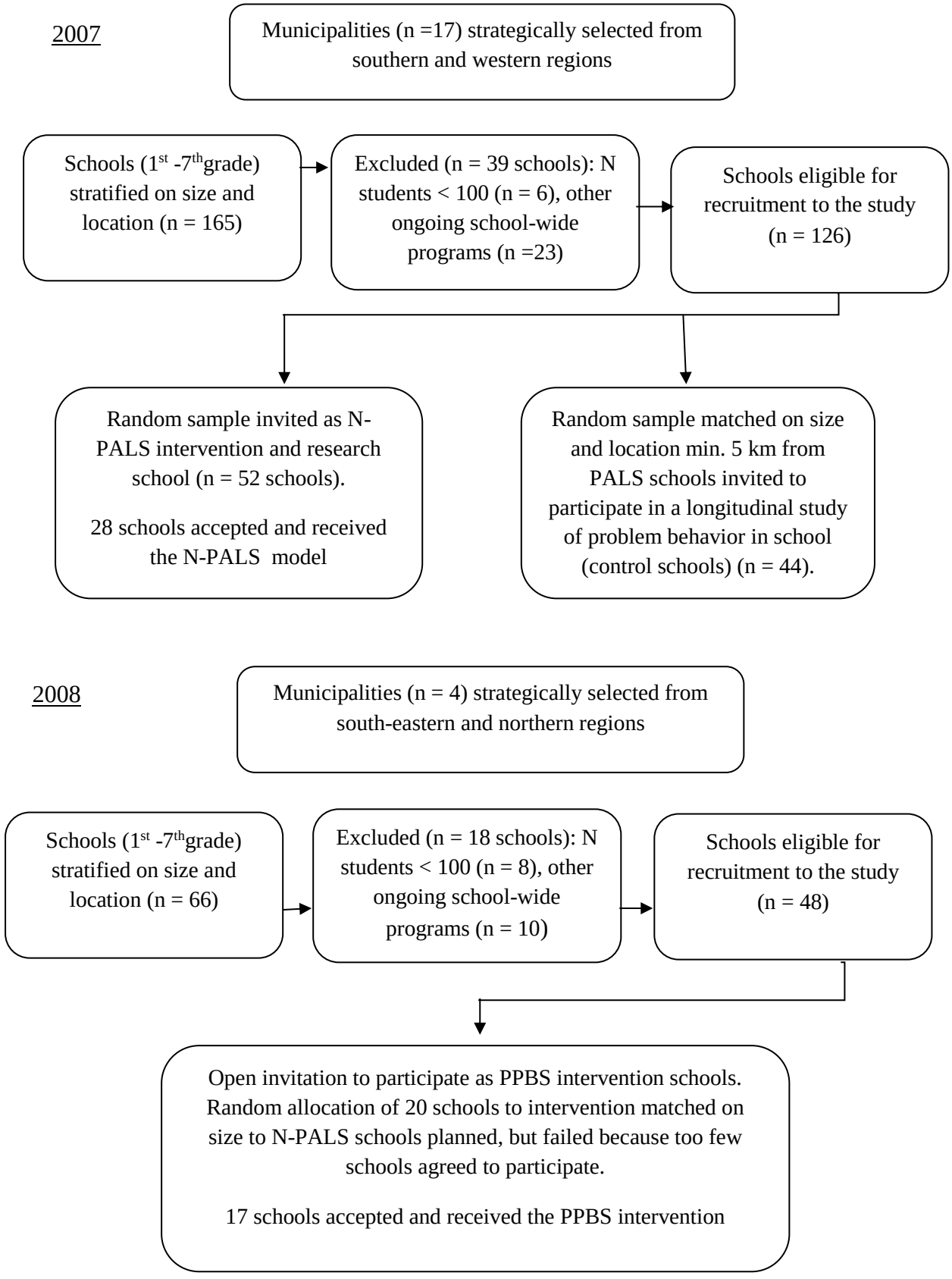
school) through a division into two sub-groups: (1) neighboring schools (located less than 5 km from a N-PALS school, i.e., a school implementing N-PALS) and (2) potential schools (located more than 5 km from a N-PALS school). To minimize the possibility of program contamination, sub-group 1 was excluded as potentials.

Step 6. A random subsample of 44 eligible schools from sub-group 2 located in the same municipalities as the N-PALS intervention group was then drawn and invited as controls (PAU, practice-as-ususal). To avoid recruiting control schools with lower perceived problem pressure or less motivation for change than the intervention schools, the schools were invited to contribute with data on the development of student behavior and school failure over time and informed that they also would serve as comparison schools. Twenty of the 44 invited schools agreed to participate (45.5%).

Step 7: Recruitment of schools to a third group (in 2008), a group that should implement an abbreviated version of N- PALS, was carried into effect by an open invitation to primary schools located in the Northern and South-Eastern parts of Norway and in different municipalities than the N-PALS and PAU groups. The invitation was advertised in relevant journals and on the NUBU´s homepage and directed to all schools situated in four strategically selected urban municipalities with no N-PALS schools. Additionally, a flyer was distributed and the local educational authorities were informed by e-mail.

Step 8: Prior to invitation, all primary schools in these four municipalities (N = 66) were, as in the other conditions, stratified into three subgroups according to school size. Eight schools were excluded from the eligible pool due to fewer than 100 students, and another 10 schools were excluded due to ongoing implementation of other school-wide programs. The eligible sample for the PPBS condition consisted of totally 48 schools (73%) that were matched to the N-PALS group based on school size. A random sample (N =20) of all schools that were positive to participate was invited to the PPBS condition. This step of the

Figure 4.1. *Recruitment flow chart* (Sørlie & Ogden, 2015, p. 6



recruitment procedure had, however, to be abandoned because too few schools agreed to participate. To approximate the desired sample size, all positives (N = 17; 25%) thus were included, regardless of school size.

4.2 Considering Study Designs

The research question together with considerations of the quality of evidence that various research designs provide, should determine the design in every study. How trustworthy, generalizable and accurate the results from intervention studies are, vary from study to study, dependent in large on the design of choice (e.g., Shadish, Cook, & Campbell, 2002; Walliman, 2006). Attaining valid effect estimates in studies conducted in complex contexts such as schools, and in studies of comprehensive interventions, such as N-PALS, are particularly challenging (e.g., Deeks et al., 2003).

Research on school-based interventions has repeatedly been judged to lack methodological rigor (e.g., Cook, 2003; Humphrey, Lendrum & Wigelsworth, 2013; Weare & Nind, 2011) due to a lack of control groups, small number of clusters (e.g., schools) or individuals (e.g., students, teachers), and inadequate descriptions of the methodological procedures. As indicated in the literature review (Chapter 3), also the early research on the SWPBS model has several methodological limitations. First, the validity of the research outcomes are often questioned due to too simple research designs. Second, key constructs are infrequently measured at the appropriate unit of analysis. For example, when individual (e.g., student) level variables are aggregated to group level (e.g., school), analyses of intervention effects, could be compromised by aggregation bias, deflated standard errors and heterogeneity of regression (Creemers & Scheerens, 1994; Welsh, 2003). Third, the statistical analyses used to explore intervention effects were inadequate or did not match the complexity of the data. For example, the significance of the observed changes in student behavior from pre to post

assessment was often not tested but described in terms of percent. Moreover, in previous quasi-experimental studies only single-level analyses were used (e.g., t-test, ANOVA, ANCOVA), and thereby ignoring nesting or clustering in data, which may have led to overestimation of intervention effects (Bickel, 2007). In order to give as clear answers as possible to the research questions of the current study and to contribute to the evidence-base of SWPBS, we tried to add up to a robust research design and an adequate analytical approach that met with the critique of prior evaluation studies and requirements of evidence-based practices (Flay et al., 2005).

Several types of descriptive (also called observational) and experimental research designs might be adequate in intervention studies, each with strengths and limitations. Farrington, Gottfredson, Sherman, and Welsh (2002) rank various evaluation designs in five groups or levels according to the validity of study outcomes. At level 1 (least valid) they place correlation studies, e.g. the correlation between a prevention program and measures at one time point of student problem behavior. Level 2 includes studies with pre-post assessment but no control group (e.g., single-case studies). Level 3, which is the lowest level any conclusion of intervention effects can be drawn, includes studies with pre-post assessments and an experimental and a control group. This design has few problems with internal validity, except for the weakness that the two groups may differ at pre-test (c.f., selection bias). Level 4 includes studies with pre-post tests, more than one experimental group in addition to a control group and control for other variables than can influence the intervention outcomes. Level 5 (most valid) includes studies with pre-post tests, two experimental groups, a control group, control for possible confounders, and randomized allocation to research condition. These designs are briefly presented and discussed in the following paragraphs, except for the Level 1 design which is not seen as adequate. Next, the rationale for the research design used in the current study is accounted for.

4.2.1 Descriptive designs. The simplest and most frequently used descriptive designs (level 2) are the *single-case* and *case-series* designs. In such studies, data on the situation (usually measured once) as ´it is' are reported respectively for one subject (e.g., one SWPBS school, one special education student) or for a few subjects (e.g., two SWPBS schools). Some SWPBS studies used a more advanced variant of the single-case/case-series design (c.f., Chapter 3) in which data collected on the situation after implementing the system-wide approach for some time (e.g., one year) was compared to already existing archival data on how the situation was prior to the intervention (e.g., the year before). Another example of descriptive designs is the *case-control* design. A well-designed case-control study can provide suggestive evidence of causal connections (Given, 2007). In such studies, subjects with a particular attribute, such as a school with high prevalence of serious conduct problems, is compared with subjects without the attribute, such as a school with no or few incidents of serious problem behaviors. Comparison is made of the exposure to something that one suspect is causing the cases, for example the fidelity with which a preventive intervention was implemented. Case-control studies are retrospective because the focus is on conditions in the past that might cause subjects to become cases rather than controls.

A third example of descriptive designs is the *cohort study*, in which a defined group of people (cohort) is prospectively or retrospectively followed over time. A prospective study follows the cohort into the future and thus can take long time to accomplish, while a retrospective study identifies subjects from past records and follows them to the present. Cohort studies provide indications of intervention effects by comparing the outcomes for subgroups of the cohort who were respectively exposed or not exposed to the intervention of interest. Both original survey data and secondary data, such as national registry data and archival school data, can be used.

As concerns the quality and generalizability of evidence that descriptive studies

provide for a cause-and-effect relationship between variables, the single-case and case-series designs are the weakest (e.g., Given, 2007; Shadish et al., 2002), while the cohort study design is among the strongest. Although results from descriptive studies cannot be used to give definitive answers on intervention effects, descriptive designs help provide answers to the questions of who, what, when, where, and how associated with a particular research problem (Given, 2007). An intervention study with some type of descriptive design can be a good starting point to decide if proceeding to a more robust evaluation design is worthwhile. Accordingly, a descriptive design was not seen as a relevant choice in order to provide valid effect estimates in the current study.

4.2.2 Experimental designs. The classic experimental design includes an experimental group and a control group. The control group is either "doing practice as usual" or receives an alternative intervention. The intervention or treatment (independent variable) is administered to the experimental group only. Both the experimental and the control group is measured at least twice on the same dependent variable (e.g., student problem behavior); once prior to or at the initiation of the intervention and once at the end of the intervention (pre-test, post-test). To evaluate if the intervention had any effects, the change in the experimental group from pre to post-test is compared with the pre-post change in the control group. One assumes that if the change in the experimental group is significantly more positive than in the control group, this is caused by the intervention. A wait list control group may also be used. A wait list control group is a group of participants who do not receive the actual or any other intervention, but who are put on a waiting list to receive the intervention after the experimental group and the research period is over.

As with descriptive designs, there are several types of simple and more advanced experimental designs (e.g., the time series design, crossover design, regression discontinuity

design, quasi-experimental design, and the randomized controlled trial design). Among the most frequently used, are the randomized controlled trial design (Level 5) and the non-equivalent comparison group design (NEC, Level 3 or 4), which often is referred to as a quasi-experimental design.

Randomized controlled trial. An experiment in which individuals are typically allocated to two groups by random (e.g., by tossing a coin) is called the randomized controlled trial design (RCT). One group receives the intervention of interest while the comparison group (control group) receives no intervention (e.g., conventional treatment) or an alternative intervention. To explore whether a given intervention can yield effects of significance for the target group, the use of comparisons is highly preferable. Comparison groups can signalize what would have happened in the absence of the intervention. The two groups are then followed prospectively and compared on measures collected at least twice; one prior to and one at the end of the intervention (pre-post assessment). In evaluating intervention effects, this design is considered the most rigorous and reliable (˙the gold standard', Shadish & Cook, 2009; Weisburd, 2010).

For testing intervention effects in educational settings, cluster randomized trials (CRTs) often are more adequate. CRTs are research designs in which clusters of individuals (e.g., classes, schools) rather than individuals are randomly assigned to two groups (or more). Consider, for example, a situation in which the researchers wish to compare the outcomes of high-frequent and low-frequent verbal praise (e.g., 8 vs 4 times per hour) for teaching expected classroom behavior. It would be difficult in practice to randomly assign students to these conditions since students are naturally ˙nested' within their classrooms. It also would be difficult for the implementers (teachers) to adjust their disciplining practices to accommodate random assignment of students. A second example is if the outcomes of a school-wide behavioral program (e.g., N-PALS) are to be compared with the outcomes of another school

behavior program or “practice-as-usual”. In such a case, schools (clusters) would be randomly assigned to implement the new intervention program or continue their "practice-as-usual."

Random allocation ensures that the intervention and control groups are similar at baseline in all respects (both on observed and non-observed variables) except of the intervention being tested. That is, the assignment of study participants is determined by chance and cannot be predicted or influenced by the researcher or by the participants themselves. Randomization helps to strengthen the probability that any observed difference between the groups at post-test is due to differences in the intervention condition alone and not effects of confounding variables (known or unknown) or selection bias. In other words, the greatest advantage of the randomized controlled trial design over any other type of design is that we can infer that a statistically significant between-group difference on the outcome can be causally attributed to the manipulated variable (the intervention).

Ethical considerations may sometimes hamper the use of random allocation, for example it would be unethical to randomize individuals to a control condition with a potentially harmful alternative intervention (e.g., exposure to high level of mental stress). Other times experiments based on the RCT design are not possible to perform due to lack of compliance by the subjects. In cases where the study subjects do not accept to participate because of the randomization procedure, a nonrandomized controlled experiment can be an alternative.

Nonrandomized controlled experiment. There are numerous variants of non-randomized designs, but the nonequivalent control group design (NEC) is the most frequently used and the variant that is closest to the RCT-design (Shadish et al., 2002). Normally, NECs use a treatment (intervention) group and an untreated control group that is strategically selected or self-selected (Shadish et al., 2002). The comparison group often consists of an opportune sample of persons or units (e.g., schools), and either active or passive controls are

used. An active control group may either offer ordinary supports or teaching (˙practice as usual') or implement an alternative intervention program. Passive controls offer no support or treatment (e.g., a wait-list group) (Shadish et al., 2002). The intervention group in NECs (as in RCTs) is compared to the control group at least prior to and immediately after the intervention. If there is no or small group differences at pre-test and the post-test scores of the comparison group differs significantly from the scores of the intervention group (pre-scores should be controlled for), the results are interpreted to indicate intervention effects.

Effects from RCTs and NECs have been compared in several meta-analyses. The conclusion is not clear. Sometimes the intervention effects observed in RCTs are larger (Shadish, Navarro, Matt, & Phillips, 2000; Shadish & Ragsdale, 1996), smaller (Kownacki & Shadish, 1999), or similar to those observed in NECs (Lipsey & Wilson, 1993). Under certain conditions, however, results from nonrandomized and randomized experiments are equally accurate (Shadish et al., 2002; Cook, Shadish, & Wong, 2008; Heinsman & Shadish, 1996; Shadish, Clark, & Steiner, 2008; Shadish & Cook, 2009; Shadish, 2011). In particular, effect estimates from NECs with high validity have repeatedly been found equal to those in RCTs (Ragsdale, 1996; Shadish et al., 2008; Shadish, Galindo, Wong, Steiner, & Cook, 2011; Steiner, Cook, Shadish, & Clark, 2010). Hence, identifying plausible threats to the validity (i.e., confounders that provide alternative explanations of intervention effects) in intervention studies based on nonrandomized designs is essential. Additionally, whenever plausible threats are identified, Shadish (2011) stresses the importance of careful incorporation of elements to the design that can strengthen the validity of the outcomes by eliminating or at least reduceing potential confounders.

In line with Farrington et al. (2002), a meta-analysis conducted by Wilson and Lipsey (2001) on the basis of 319 school-based intervention studies, concluded that randomized experiments and quasi-experiments are the most appropriate and those that give the most valid

outcomes. However, the use of controlled experiments (randomize and non-randomized) in evaluations of school-based interventions is rare, not only in Norway and Sweden (e.g., Eklund, 2000) but world-wide. For example, Hsieh et al. (2005) documented a general decline in both randomized and non-randomized educational intervention studies from 52% in 1983 to 33% in 2004. In a later overview, Sanders (2011) found that of 476 educational research articles published in four selected journals between 2007 and 2009, only 16% (N = 75) employed randomization, a finding consistent with Shaw, Walls, Dacy, Levin, and Robinson (2010) who examined a selection of 243 articles in the Web of Science database from 1995 to 2005. In a Swedish overview, only six percent of educational dissertations between 1997 and 2006 evaluated intervention effects, of which only one used randomization (Sundell & Stensson, 2010). The lack of school-based studies with an experimental design may be due to that such studies are complicated to implement, time consuming, and expensive to carry out. Another reason may reflect that methodological questions within the educational disciplines "most often are strongly loaded, both ideologically and politically" (authors translation of "oftast starkt ideologiskt-politiskt laddade", Eklund, 2000, p. 138), and that this may negatively influence the schools' willingness to participate in quantitative experimental studies as well as the school political authorities' willingness to finance controlled studies on pre-structured intervention programs.

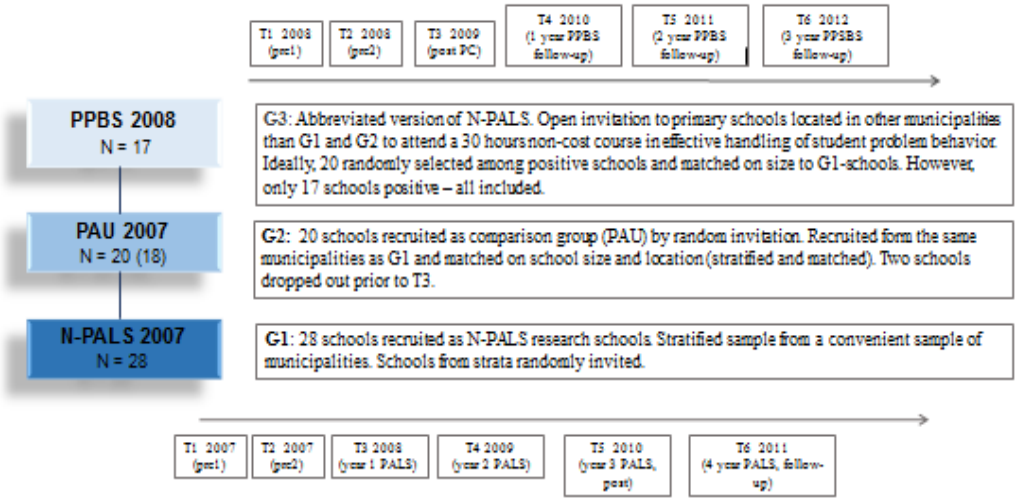
4.3 The Design of the Current Study

After considering several possible ways of designing the study, a two-group longitudinal randomized controlled trial (RCT) was, as earlier mentioned, considered the most appropriate design, and the one that best could contribute with valid estimates of the N-PALS intervention effects. However, the schools' unwillingness to participate on the condition of randomization made us change to a nonrandomized study design. Fortunately, re-inviting the schools to an

alternative study approach turned out successfully. After considering several alternatives, a decision was made to utilize a strengthened nonequivalent control group design (NEC).

The strengthened study design corresponds to level 4 in Farrington et al.'s (2002) ranking of evaluation designs, and includes three ordinary primary school groups (Grades 1-7); one group of 28 schools implementing the full scale model of N-PALS, a second group of 20 control schools doing ˙practices-as-usual' (PAU), and a third group of 17 schools implementing the abbreviated PPBS intervention (Figure 4.2).

Figure 4.2. *The study design*



An open cohort design was applied, which implied that staff members and students new to the schools and new cohorts of 4th graders could be enrolled in the study at each measure point, except for at T5 and T6 in that at data from at least two time points were desirable of analytic reasons. Due to a lack of resources, staff members and students who moved or left the participating schools (e.g., by the end of 7th grade) during the study period were not followed.

In modelling the study design, we were aware that quasi-experiments frequently misestimate effects (Heinsman & Shadish, 1996; Shadish et al., 2002; Shadish & Ragsdale, 1996). Accordingly, and as described in Paper 1 and further elaborated on in the following

paragraphs, several elements were added to the NEC design in order to strengthen the validity of our findings.

4.4 Threats to Validity in Evaluation Studies

Several reviews of validity threats across nonrandomized designs have been published, of which the first was by Cook and Campbell in 1979 and later by Shadish et al. (2002), who also provided advices on how to neutralize these threats. According to Shadish et al. (2002), inferences in quasi-experiments like ours can be improved by adding certain elements to the design, such as more than one pretest assessment, use of stratification and matching procedures and multiple control groups. Specific statistical methods can also be used in the analytic stage (e.g., controlling for pretest scores, adding covariates related to the selection process). Shadish (2011, p. 642) and Cook and Steiner (2010, p. 57) stress, however, that "There is no substitute for good prospective designs of nonrandomized experiments.", and "…it is not possible to put right with statistics what has been done wrong by design".

The eight most serious threats to internal validity are; 1) history, 2) maturation, 3) testing, 4) instrumentation, 5) selection, 6) regression to the mean, 7) attrition, and 8) contamination.

1) *History* refers to intervention-independent events occurring between pretest and posttest (e.g., change of principal, turnover in the school staff, changes in socio-economic conditions) which can exert an influence on the individual's performance or ratings and thereby affect the estimation of intervention effects (e.g., spuriously low).

2) *Maturation* refers to naturally occurring changes in characteristics of the study participants (e.g., physical, intellectual, emotional). Such changes can easily, although incorrectly, be interpreted as impacts of the intervention.

3) *Testing* relates to that a pretest (e.g., completing a questionnaire) might increase or

decrease the study participants' sensitivity or responsiveness to the experimental variable. That is, exposure to a pretest may, independently of the intervention, influence the performance on subsequent tests. Observed improvements from pre- to posttest may thus partly be due to testing rather than to the intervention. Likewise, may program training affect the participants' awareness or sensitivity of the problem in focus. Increased awareness may result in higher problem scores at posttest than at pretest (e.g., inflated reports of bullying), which in turn may lead the researcher to falsely conclude with negative or no intervention effects.

4) *Instrumentation* occurs when the researcher changes or refines the measurement instruments or tests from one assessment point to another. Observed changes in the participants' scores from pretest to posttest may then reflect changes in the instruments rather than changes following the intervention. That is, validity threats stemming from instrumentation may produce illusory intervention effects.

5) *Selection* refers to bias in how subjects or clusters are selected to a study or the way they are assigned to the intervention and control group. Selections bias may result in groups that systematically hold different characteristics prior to the intervention or in a research sample that is not representative of the target population. Selection bias frequently occurs when nonrandom procedures are used or when random assignment fails to balance out differences among subjects or clusters across the different experimental conditions. Selection bias is seen as the primary threat to internal validity in nonrandomized experiments (Deeks et al., 2003; Shadish et al., 2002; Ukoumunne et al., 1999; Weisburd, 2010).

Initial group differences may affect the outcome estimates because if the scores of the intervention and control groups are different at the end of the experiment, this may be for reasons other than the intervention. If, for example, only those who will be most involved in the intervention are invited as the sole informants, some might provide biased assessments

due to the so-called Hawthorne Effect or the Novelty Effect. The Hawthorne Effect (Landsberger, 1958) refers to a process where the participants in an experiment subconsciously change their behavior or post-ratings simply because they are being studied (feel important or chosen). The Novelty Effect (Gravetter & Foranzo, 2011) refers to the effect of being introduced to a new intervention or innovative treatment. If participants have strong or low beliefs in the intervention or treatment program or strongly trust or distrust the implementer(s) of the intervention, this may color their ratings in positive or negative ways, and thereby bias the intervention effects. In a review of educational research, Clark and Sugrue (1991) found that uncontrolled novelty effects may cause 50% of a standard deviation (SD) score increase for up to 4 weeks, decaying to 30% of a SD for 5-8 weeks and to 20% of a SD for over 8 weeks. Likewise, the John Henry Effect (Saretsky, 1972) can skew the outcome estimates. This effect occurs if the control group views itself as being in competition with the experimental group. The group then may change their behavior in beneficial ways or provide more positive or flattering ratings than are justified

6) *Regression* bias may occur when participants self-select to the study or the intervention because they have fewer or greater problems than average. When units or individuals with extreme scores are invited or selected, there will be a tendency towards less extreme scores closer to the average at post-test, while units or individuals with low scores on the pre-test tend to be higher on the post-test. Essentially, no change has taken place due to the intervention. This phenomenon is often referred to as ˙regression to the mean' and can confound estimates of intervention effects (Cook & Campbell, 1979).

7) *Attrition* implies that loss of respondents to intervention or assessment can produce artefactual effects if the loss is systematically correlated with conditions. Low or decreasing response rates across measurement points are akin to attrition. Hence, if high attrition or low response rates are prominent, one cannot expect to obtain good effect estimates (Jüni, Altman,

& Egger, 2001). Observed differences between the intervention and the comparison group at posttest may then be due to the remaining non-representative participants rather than to effects of the intervention (Shadish et al., 2002). That is, attrition is a threat to validity if those who drop out of an experimental study are significantly different from those who remain in the study.

8) *Contamination* (sometimes referred to as diffusion) is a validity threat that occurs when those who are not intended to receive an intervention inadvertently do so, for example, if schools in the control group learn about and implement significant parts of the intervention under evaluation (Keogh-Brown et al., 2007). This type of contamination tends to confound the effect estimates. The chance of attaining non-significant effect estimates increases if contamination is present.

Except for the validity threat stemming from selection bias, the other validity threats listed, are equally likely in nonrandomized and randomized experiments.

4.5 Strategies to Reduce Threats to Validity in the Current Study

In this paragraph, strategies to reduce the most likely threats to the validity in the current study are accounted for.

4.5.1 Efforts to minimize selection bias. In order to minimize threats to validity stemming from selection bias in nonrandomized studies, Shadish and his colleges (2002) recommend that the following four elements should be added to the design; 1) use of multiple non-equivalent comparison groups rather than just one, preferably a local or internal control group in addition to an external control group, 2) use of active rather than passive controls, 3) use of stratification and matching procedures prior to group allocation, and 4) use of distinct procedures to reduce self-selection. All four design elements were added to the current NEC-

design.

Use of multiple and active comparison groups. In line with the recommendations, a three-group comparison design was preferred in the current study. Selection bias would occur if the participants in the N-PALS group differed significantly from those allocated to the alternative intervention (PPBS) and the practice-as-usual group (PAU). The inclusion of both an alternative intervention and a control group might reduce bias and putting the N-PALS model on a much stronger test than in the first effectiveness study (Sørlie & Ogden, 2007; Ogden, Sørlie, & Hagen, 2007). To achieve more accurate results and realistic effect estimates, we decided to operate with both a local or internal comparison group and an external comparison group (Shadish et al., 2002; Ukoumunne et al., 1999) as well as with active rather than passive comparisons (Heinsman & Shadish, 1996). Shadish (2011, p. 640) emphasizes that it is important that the controls should come from the same location as the intervention group, share as many observed substantive characteristics as possible at baseline, and that "the goal of a focal local control is to reduce as much of the difference between treatment and control as possible in the design phase of the study so that there is less bias to remove at the analytic stage". Compared to active controls, passive controls (e.g., wait-list groups) have been found to generally yield larger intervention effects. That is, withholding active interventions from the control participants will normally result in a treatment-control discrepancy that is larger than when controls receive active interventions (Heinsman & Shadish, 1996).

In the current study, the local controls (PAU) thus were drawn from the same population as the N-PALS group, while the external controls (PPBS) were drawn from a different population. The PAU group served as active local controls, while the PPBS group served as active external controls.

Use of stratification and matching procedures prior to group allocation. Stratifying

implies that units are sorted in homogenous sets (strata) that contain more units than the experiment has conditions (Cochran, 1968). Matching implies that units with similar scores on the matching variable are grouped (or paired) so that the intervention and comparison group "each contains units with the same characteristics on the matching variable" (Shadish et al. 2002, p. 118). Different from stratification, matching contains the same number of units as conditions. Shadish et al. (2002) recommend that assignment of units from strata or matches should be used both in randomized and non-randomized studies whenever feasible and a good matching variable can be found. Moreover, the benefits of stratification and matching are largest if the methods are used prior to assigning units to the research conditions (Shadish et al., 2002). Optimally, each intervention unit should have multiple matched controls and vice versa (e.g., Cochran, 1983). The more variables that are used to increase the likelihood initial group similarity, the more difficult the matching becomes.

Consequently in the current study, a stratifying and matching procedure based on school size and location was decided to increase the likelihood that the comparison group would resemble the N-PALS group on key characteristics (for details, see Recruting schools to the study). School size was used as stratification variable due expected relations both to initial school variation in behavior problems, intervention outcomes, and implementation quality (program fidelity, adherence). For example, the majority of previous studies of school size and student behavior indicate that smaller schools (less than 300 students) are associated with slightly less behavior problems, more safe and orderly environments and higher student achievement than in larger schools. The effects of school size on student behavior may, however, be more indirect than direct. For example, Pittman and Haughwout (1987) found that the relationship was mediated by school climate. Using hierarchical linear modeling in a study of a national sample of 2,213 male public high school students in 10th grade, drawn from 87 randomly selected schools, Brezina, Piquero, and Mazerolle. (2001) found that

school size correlated in the lower range with aggression (r =.20-.23) but exhibited significant effects on both aggressive behavior and conflict with peers.

In addition to our attempts to reduce selection bias, efforts were made to minimalize other plausible threats to the validity by adding more elements to the NEC design, such as multiple assessment points, multiple informant groups, careful consideration of which measures to use, and planned strategies to minimize attrition and low response rates.

4.5.2 Use of multiple assessment points. Beyond the standard pre-post tests, recommendations have been made to include more measurement points both prior to, under and after the intervention in order to reduce validity threats stemming from study testing, maturation, attrition, and regression artifacts (Shadish et al., 2002; Ukoumunne et al., 1999). Consequently, a total of six measurement points (T1-T6) were decided in the current study, including a double pre-test, two intermediate tests, a post-test, and a one-year follow-up test. T1 (baseline) and T2 were conducted, respectively in April and August prior to and parallel to the initiation of the interventions, while T3-T6 were conducted in June the three subsequent years.

4.5.3 Use of a multi-informant approach. To remove the risk of mono-informant and testing bias, multiple informant groups were used in the current study. Together with the fact that there is no 'gold standard' informant group that is expected to provide more valid and reliable data (Kraemer et al., 2003), the broad study perspective of the current study called for a multi-informant approach by including the perspectives of principals, teachers, assistants, after-school personnel, parents, and students. Multi-informant variability may, however, lead to incongruous study conclusions if not handled with caution in the analyses of intervention effects and in the interpretation of results (Horton, Roberts, Ryan, Suglia, & Wright, 2008;

Kraemer et al., 2003). When multiple informants contribute with information regarding a target characteristic or construct, data convergence is almost never achieved. More than moderate inter-rater agreement is rare and should not be expected in the present study. For example, correlation estimates between teacher- and parent-assessed externalizing and internalizing behavior tend to lie at approximately .13 and .32, respectively, while the correlation between children's and adults' reports seldom exceeds .20 (Kraemer et al., 2003). Corresponding informant discrepancies have also been reported in intervention studies (e.g., De Los Reyes, 2011).

4.5.4 Use of relevant and reliable measures. In order to eliminate or at least reduce threats to the internal validity tied to instrumentation, a well-established measurement battery was central in the design of the N-PALS evaluation study. As noted by Ogden (2010), in experimental studies the researchers need to have hypothesizes of which outcome variables the intervention should primarily influence upon (e.g., externalizing student problem behavior in N-PALS) and how to measure these. In addition to operating with primary and proximal (immediate) outcome variables, more distal (longitudinal) and secondary outcome variables (e.g., unintended impacts on internalizing problem behavior) can be relevant. Ideally, measures should have been proven reliable and valid in prior studies, i.e., the degree to which the measures capture what they intend to measure and have acceptable psychometric properties (e.g., high internal consistency, predictive validity, and sensitivity to growth) (APA, 2009).

Internal consistency or reliability is about the homogeneity of a psychometric instrument (scale), that is, if the items or questions in a rating scale or test address the construct(s) or dimension(s) of interest (e.g., externalizing and/or internalizing problem behavior), and defines the consistency of the results. A measuring instrument with high

internal reliability measures the same construct or dimension with several items and several answering values (e.g., 1-5, 1 = never, 5 = daily). A measuring instrument is said to be reliable if it can produce similar results if used again in similar samples and contexts (e.g., Grazziano & Raulin, 2007). Several techniques can be used to assess the reliability of a measure instrument (e.g., split-half test, test-retest, interrater test). One of the most frequently used, is the Cronbach's alpha (α, Cronbach, 1951). Cronbach's alpha is a function of the number of items in a scale or test, the average covariance between items, and the variance of the total score. The alpha coefficient ranges in value from 0 to 1, with 0.7 generally accepted as a sign of acceptable reliability. If alpha is high (e.g., 0.95), it may suggest high level of item redundancy; e.g., that a number of items ask the same question in slightly different ways.

Predictive validity addresses the question ʻCan the assessment results be used to predict things about the participants?' Predictive validity may for example refer to the ability of a measure to predict future behavior (e.g., criminality), performance (e.g., in math) or scores on a criterion measure (e.g., referred to special education). Predictive validity can be determined by calculating the correlation between assessment results and the subsequent targeted behavior. The stronger the correlation is, the higher is the degree of predictive validity. Sensitivity to growth refers to the amount of change that can be detected by a measurement (e.g., Lannie, Codding, McDougal, & Meier, 2010). A measure instrument might be sensitive to capturing individual differences at a given time point but not necessary equally able to catch individual changes over time. The sensitivity of a measure also includes its ability to correctly identify the proportion of individuals in a sample or population with a particular disease or condition, while its specificity refers to the ability to measure the percentage of healthy people who are correctly identified as not having the disease or condition.

4.5.5 Study variables. In the current study, the study variables were selected to match the expected intervention outcomes and to allow for examination of factors and circumstances that might moderate or confound the outcomes. Emphasis was also put on selecting instruments that allowed for measuring potential intervention outcomes at the adequate level of analysis (e.g., school level). Measures in English were independently translated and back-translated by two senior researchers at NUBU. When we were not able to find adequate and reliable measures, new were developed, piloted, refined according to factor-analyses, and reliably-tested prior to use.

In line with the goals of the N-PALS model (full and abbreviated version), our main focus was, as reflected in the dissertation, on measuring the level of externalizing student problem behavior across time, on changes in classroom climate (psycho-social learning conditions), and on changes in the school staff's disciplinary practices and collective efforts to prevent school misconduct and exclusion. Measures of possible moderators such as school size and implementation quality and possible confounders such as exposure to other programs during the study period were also included. All measures were held constant across assessment points. The specific measures applied in Paper 1-5 are described in Chapter 5.

To demonstrate our approach in selecting adequate study measures, an example follows; if the SWPBS model was to be effective within the Norwegian context, we expected positive changes in the extent of student problem behavior to occur, both within and outside the classroom context. A search in relevant research literature and databases revealed that most reliable and well-established instruments measure problem behavior at the individual level, while few existed that matched our needs of instruments to measure student problem at higher levels. However, two measurement scales developed by Grey and Sime (1989) to tap the occurrence of negative behavioral incidents as observed by teachers respectively in classrooms and on common school areas (e.g., corridors, playground, toilets) were seen as

appropriate. Both scales had previously been translated and used in several Norwegian studies with satisfactory high internal consistency (e.g., Ogden, 1995; Kjøbli & Sørlie, 2008, Sørlie & Ogden, 2007).

4.5.6 Efforts to minimize attrition and low response rates. In order to prevent validity threats due to biased dropout and promote high response rates, several strategies were added in the design phase of the study (Wineman & Durand, 1992). For example, to facilitate the implementation of the six assessment points and the communication between the schools and the research team, all schools appointed a research contact. These contact persons were trained (for one day) and individually supported throughout the study by a research coordinator at NUBU. Each year the data collections were accomplished within a time frame set by the research team. The schools could, however, decide which days within this time frame that were most adequate.

Due to the large dimensions of the study, questionnaires were seen as the most feasible data-collection method. Five questionnaires were developed; one for the principals who participated at T1 and T5 only, one for all staff members, one for the class head teachers (individual ratings of the students in their class), one for the parents (ca. 30% of sample), and one for the students in 4^{th} to 7^{th} grade.

Another example; to motivate continued participation, the schools were each year offered a small compensation for their extra work, such as free entry for five persons to the NUBU's national conference, staff candy or fruit boxes, free textbooks on behavior problems and classroom management, and an individual school evaluation report based on baseline data. The parents who completed the questionnaire participated in a drawing of 2-5 gift coupons of 2,000 or 5,000 NOK at each data point. Moreover, carefully selected preliminary results were each year presented at the NUBU's national conference and the schools were

informed as new research articles were published. For more examples, see Procedures.

4.6 Statistical Power, Effect Size and Sample Size

A challenge in studies like ours where schools rather than individuals (students, teachers) are the main unit of analysis, is to calculate *a priori* (in beforehand) how many clusters or units should be included in the study to obtain sufficient statistical power to reliably analyze intervention effects. A priori power analysis is also important of economic reasons, because the larger research studies (high N) are, the more costly they become in terms of money, effort, and other resources. At the same time, one should not throw away money by conducting studies that are not strong enough to establish valid information of intervention effects.

Statistical power decides if the study results are transferable to or representative for a greater population (c.f., generalizability) than the study sample and if a study is sufficiently extensive to demonstrate effects of the intervention (if effects truly exist). The power can be expressed in a number or percentage which indicates the probability that a study will obtain a statistically significant effect. For example, a power of 80% (or 0.8) means that a study (when conducted repeatedly over time) is likely to produce a statistically significant result in 8 out of 10 times. Low power implies less chance of detecting a true intervention effect and also gives less likelihood of the significant findings representing a true effect.

More technically, statistical power is the probability that a statistical analysis will be able to catch false null hypotheses. Statistical power is inversely related to beta and in short, power can be expressed as $1 - \beta$. In general, larger studies are more representative of the population than smaller studies, and the greater the statistical power is, the more trustable are the study results. That is, the effect estimates become more precise and the risks of Type-I error (concluding there is a significant effect when, in fact, there is none) and Type-II error

(concluding there is no significant effect when, in fact, there is one) are less in studies with high than low statistical power (Shadish et al., 2002). Consequently, low statistical power represents a threat both to the internal and external validity of study results.

Statistical power may be affected by several factors, especially the statistical significance criterion used, the magnitude of the effects of interest, and the sample size (Cohen, 1977). Large samples offer greater test sensitivity than small samples, larger effects are easier to detect than smaller effects, and the higher intervention effect the higher statistical power. Larger sample size reduce distribution in scores (i.e., lower standard deviation), leading to increased power (e.g., Maxwell, Kelly, & Rausch, 2008). The most commonly used significance criteria are $p = 0.05$ (5%, 1 in 20), $p = 0.01$ (1%, 1 in 100), and $p = 0.001$ (0.1%, 1 in 1000). If the criterion is 0.05, the probability of a pseudo effect at least as large as the observed intervention effect is 5% or less (Rossi, Lipsey, & Freeman, 2004).

4.6.1 Effect size. A significant finding (low p-value) does not, however, tell if the observed group difference is of practical interest. Particularly in large studies (high N), analyses may result in significant outcomes that are so small that they in fact are practically irrelevant. Hence, it is important in beforehand to consider how large the expected effects should be, and also later to estimate the magnitude of the observed group differences in outcomes (effect size, ES). The magnitude of the intervention effect is usually quantified in terms of a standardized effect size (e.g., Cohen's *d*; Hedge's *g*; Glass delta, Δ; eta squared, η^2), of which Cohen's *d* is the most commonly used. The formula is $d = Y - X/s$, where Y-X refers to the difference between the means for the intervention and control group, divided by the common standard deviation (SD) of the outcomes in the two groups or by the pooled SD.

When it comes to studies with multilevel designs and studies with more than two time points such as ours, there is, however, no consensus on how to calculate effect size measures

(Publications APA, 2008). We decided that an approach suggested by Feingold (2013) to guide the calculation of effect sizes in longitudinal multilevel studies was the most appropriate. Cohen's *d* was thus calculated according to Feingold's (2013) recommendations for designs with two independent groups and three or more time points (linear models). The formula is: $Y = a + b_1\text{Group} + b_2\text{Time} + b_3\text{Group} * \text{Time} + e$, where "*Y* is a mean of one of two groups at each of four times, Group is the condition coding (1/2 for control and 1/2 for treatment), Time is the time coding, Group * Time is the Group X Time cross-products, and *e* is zero (because the linear trend of the means is perfect for both conditions in the illustrative data)" (Feingold, 2013, p. 115). Then the effect size would be calculated using *b*3 in the next equation: $d = (b_3 * \text{duration}) / SD$. Feingold exemplifies (2013, p. 115): "When the number of weeks is 3 (as with three posttest scores in a study with weekly assessments) and the within-groups standard deviation is, say, 10, $d = (1)(3)/10 = 0.30$".

A general guideline is that an ES of 0.2 and below refers to a small effect, 0.5 to a medium effect, and 0.8 and above to a large effect (Cohen, 1977). Expressed more substantially, a Cohen's *d* (which is independent of sample size) of for example 0.2 implies that 58% of the intervention group will have scores above the mean of the control group by posttest. However, as commented on by Prentice and Miller (2002) and Westen, Novotny, and Thompson-Brenner (2004) even small effects can sometimes be impressive; for example when an intervention effect holds even under very challenging conditions, when only minor manipulation of the independent variable produce important changes, or when a large effect is observed for a small part of a sample (e.g., 20%) as compared to an intervention with a small effect for most (e.g., 90%). It may also be that a small effect where the unit of intervention and analysis is the cluster (e.g., school) can both be more difficult to achieve and perhaps have more far-reaching consequences in the long run than an equal effect size in a study where the unit of intervention and analysis is the individual. It may also be that a small

longitudinal effect (e.g., observed three years after pre-test) is more impressive and of greater practical significance than a small immediate effect (e.g., observed six months after pre-test).

The mean intervention effect size does not, however, reflect the existing variations in outcomes, i.e., that some participants or units may change a lot across time while others may change little. Therefore, APA (2009, p. 24) strongly recommends that the confidence intervals (CI) of the preferably standardized effect sizes always should be reported. Confidence intervals indicate both the precision of the effect estimate and its uncertainty. More precisely, CI refers to the probability that a value will fall between an upper and lower bound of a probability distribution, and are typically stated at the 95% or 99% confidence level (Field, 2013). Assume that the lower and upper intervals of the estimated d = 0.5 lies between 0.4 and 0.6, one can expect that the "true" mean intervention effect will lie within this 'uncertainty range' in 95 out of 100 similar studies. In the current study, the upper and lower confidence intervals were calculated using Grissom and Kim's formula (2005, p. 59-60):

$$\text{Sd2} = \frac{na+nb}{nanb} + \frac{d2}{2(na+nb)}$$

4.6.2 Sample size estimation. Based on results from the prior study of the N-PALS model (Sørlie & Ogden, 2007), the initial standard sample size calculation with an effect size (d) of .30-49, two independent variables, p = .05 and β = .80 (power), indicated that 23 - 35 schools should be included. However, as noted by Hedges and Rohads (2010, p.1), the "relationships among units within clusters usually implies greater similarity of outcomes for units within a cluster than for units coming from different clusters". That is, standard sample size calculation will lead to sample sizes that are too small because clustering impacts the statistical power. To adjust for clustering in the present study the intraclass correlation coefficient (ICC, p) and the Design Effect were used (Ukoumunne et al., 1999). The ICC was defined as the portion of the total variation in outcome that can be attributed to differences

between the clusters. In other words, ICC tells the proportion of the total variance in Y that is accounted for by the clustering (nesting). The formula used was, $\rho = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_e^2}$ where σ^2_s is the between-cluster variance component and σ^2_e is the within-cluster variance component.

Hedges and Hedberg (2007) report that values for ICC on a variety of studies of school performance in American schools are typically in the .10 to .25 range, while values for measures of nonperformance (e.g., problem behavior, social skills) are typically more modest. Generally, Norwegian schools are less different from each other than American schools (Ogden, 2014), so lower ICCs should be expected. For example, ICCs below .05 are observed in studies of bullying (e.g., Idsøe, Dyregrov, & Idsøe, 2012).

The Design Effect refers to how many more subjects a cluster-based study should include as compared to a study in which individuals are randomized. The formula used was: Deff = 1 + (n-1) *p*, where n is the average cluster size and *p* is the intra-class correlation coefficient of the outcome. The estimated ICC (0.056) for school problem behavior in the pilot study (Sørlie & Ogden, 2007) said that the number of students in the current study should be at least 10,083 (N = 735 x 13.7) to achieve sufficient statistical power and the adjusted sample size estimation indicated that between 64 and 76 schools should be included (calculated on the basis of an average school size of 245 c.f. Norwegian standard). The a priori calculation fitted well with results from simulation studies which have shown that fewer than 50 groups can lead to biased estimates of the standard errors at the group level (Maas & Hox, 2005). An operational sample size of 65 schools was decided on practical and economic reasons.

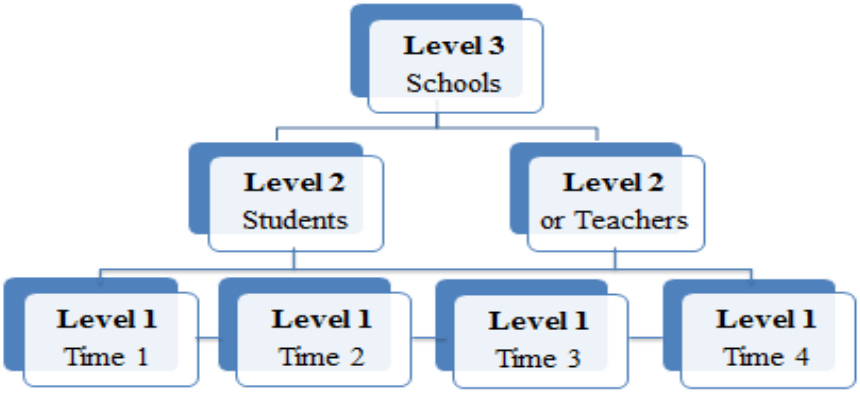
4.7 Analytic Approach

As in most educational research, the data in the current study had a hierarchical and nested structure which required multilevel modeling analysis to avoid Type-1 errors (Bickel, 2007;

Field, 2009; Heck, Thomas, & Tabata, 2010; Hox, 2010). The concepts of multilevel models or multilevel analyses are also known as hierarchical linear models, mixed-effects models, random effects models or random coefficient models, and variance component models. Nesting in data occurs when people are sampled from similar contexts, for example, when data are collected from students within classrooms or schools, and when data are collected on multiple occasions over time (repeated measures).

To evaluate the effects of the N-PALS model and the PBPS intervention, a three-level model was applied, as illustrated in Figure 4.3 (level 1 = time, level 2 = students/teachers, level 3 = schools) and specified in Paper 2-4. To investigate the concurrent and prospective relationships between teacher collective efficacy and problem behavior in school, a two-level model (level 1 = teachers, level 2 = schools) was applied, as specified in Paper 5.

Figure 4.3. *Example of a hierarchical or nested data structure*



The underlying theory of nesting and multilevel analysis is that ”individuals interact with the social groups or contexts to which they belong, and that those groups are in turn influenced by the individuals who make up the group” (Hox, 2010, p. 1). Consequently, the students’ behavior and motivation in one class can be more similar to each other than to the

students' behavior and motivation in another class because of different classroom experiences and/or because of differences in the teachers disciplining and teaching practices. It can be applied, that students in one school often will be more similar to each other than to students in other schools, this because schools may differ in their policy and tend to reflect their socio-demography (which may differ from school to school).

The individuals (e.g., students) and social groups (e.g., classes) or contexts (e.g. schools) can be "conceptualized as a hierarchical system of individuals nested within groups, with individuals and groups defined at separate levels of this hierarchical system" (Hox, 2010, p.1). Multilevel models are developed for analyzing such hierarchically structured data and two or three-level models are most common. For example, in school-based studies data are often collected from many students. The students then make up the lowest level in the hierarchy (level 1) while the classes to which the students belong can be seen as a level up in this hierarchy (level 2), and the classes can be seen as nested within schools (level 3). We then operate with a three-level model. Another example (c.f., Figure 4.3) is when data are collected at several time points (level 1) from several students and/or teachers (level 2) in several schools (level 3). Data collected from a student or teacher at several successive time points are most probably not independent of each other. The repeated measurements can thus be said to be nested within students/teachers which themselves are nested within schools.

Multilevel analyses have since the 1980's become a standard method in school effectiveness research (Hox, 2010; Peugh, 2010). However, in analyzing school intervention outcome data this statistical approach has been used surprisingly seldom given the frequent hierarchical structure of the data. More conventional statistical methods, such as variance and covariance analyses (e.g., *t*-tests, ANOVA, ANCOVA) and regression analyses (e.g., ordinary least square regression, multiple regression), are instead often used to analyze school intervention effects, even though these methods lean heavily on the assumption of

independent observations. For example, in analysis of covariance one assumes that "the relationship between the outcome and covariate is the same across different groups that make up the predictor variable" (Field, 2009, p. 729). Nesting implies that independency in the data is unlikely to be true. The assumption of independence in the data maintains that each score within a group is independent of other scores. That is, one assumes that each score within a particular group contributes with unique information equal in weight to the other scores in that group (Peugh, 2010). Ignoring nesting or dependency in the data may form a threat to the validity of the statistical inference in school-based effectiveness studies and lead to biased *p*-values, incorrect confidence intervals, and inflated effect sizes (Baldwin, Murray, & Shadish, 2005; Wampold & Serlin, 2000). In such cases, the probability of incorrectly concluding that an effect is statistically significant can be substantially higher than the commonly accepted lowest significance level ($p = 0.05$). Aarts, Verhage, Veenvliet, Dolan, and van der Sluis (2014) concluded their findings based on a simulation study that ignoring nesting can result in a Type-1 error rate as high as 0.80. This means that if no intervention effect is present, conventional methods as those mentioned above, may yield spurious significant intervention effects in 80% of the studies (standard errors are much too small).

Hence, when the structure of data is nested or hierarchical, standard analytic techniques are inappropriate for estimating linear models, and multilevel analysis should be used (Raudenbusch & Bryk, 2002). Multilevel models recognize the existence of such data hierarchies by accommodating the correlation among observations and by allowing for residual components at each level in the hierarchy. The within- and between-cluster variability is respectively modeled via an additional error term called random effect (Raudenbush & Bryk, 2002; Singer & Willet, 2003). For example, a two-level model which allows for grouping of student outcomes within schools would include residuals both at the student level and the school level. In multilevel analyses the residual variance is split up into a

between-school (cluster) component (i.e., the variance of the school-level residuals) and a within-school component (i.e., the variance of the student-level residuals). The school-level residuals (often referred to as ˙school effects'), represent unobserved school characteristics that influence the students outcomes. By estimating such shared variance (within-school, between-school) improved accuracy of the intervention effect estimates is achieved (Raudenbusch & Bryk, 2002).

In addition to running the risk of Type-I error inflation and misestimated standard errors, ignoring nesting and hierarchy in data can lead to ecological fallacy (Robinson, 1950). Ecological fallacy refers to the erroneous assumption that individuals have the salient or average characteristics of the social group he or she belongs to. In single-level modeling, such fallacy occurs when inferences made at the student level (level 1) are mistakenly applied to the school level (level 2), or vice versa. However, by partitioning the variance multilevel analysis addresses this issue. Additionally, while single-level models (e.g., ANOVA) increases the degrees of freedom to be used and reduces the size of standard errors, multilevel models adjust for measurement error at level 1 (e.g. student level).

Other benefits by choosing multilevel analysis as the main analytical approach in the current study are that multilevel models can accommodate longitudinal data where participants are measured on different schedules, treat time as a as categorical variable, continuous variable, or some combination, permit heterogeneity of regression across schools by estimating parameter values for each school and modeling the variation in coefficients between schools, and also better handle unbalanced group sizes (Hox, 2010, Singer & Willet, 2003). Multilevel data can also be analyzed using ANOVA models for nested and repeated measures designs. However, as noted by Hox (2010), complications occur in studies where the group sizes are unequal and/or the time between repeated measures vary. In our case, the time span between assessment point one and two (T1, T2, ca. 4 months) differed from the

time span between the other assessment points (ca. 10 months), and the size of the three research groups (N-PALS, PPBS, PAU) was unequal. Multilevel models are also designed to solve the problem of cross-level interactions, for example how several student and school variables jointly influence a particular student outcome. Another advantage of multilevel models is that the individual parameter estimates are based on data from all cases and not only from ˙completers'. The estimates can therefore be relatively reliable even with a small number of observations per case and more accurate when data are missing (e.g., Atkins, 2005; Little & Rubin, 2002; Schafer & Graham, 2002).

4.8 Ethical Aspects

The standards of the ethical review board ˙Regional komité for medisinsk forskningsetikk – Sør-Norge' (Regional committee for medical research ethics, Norway South) and of the ˙Norsk senter for forskningsdata' (NSD, Norwegian Social Science Data Services) were followed throughout the conduct of the study. The Declaration of Helsinki (Helsinki, 2010) requires informed consent form participants in research studies. For children and others with impaired consent, the declaration allows for consent to be given by others, like parents or other guardians (e.g., foster parents). However, the child can independently decline participation. In line with this, all parents and students were in writing informed of the purpose of the study, who was responsible, what participation would imply, and that they could withdraw from the study and ask for their data to being deleted at any time point and without explanations. The schools were also verbally informed of the study and their individual right to privacy when participating in research on regionally organized meetings. The class head teachers were responsible for distributing the information letter to the parents and students and for collecting the parental consents prior to data collection. Consent from the teachers was collected in parallel with their completion of the questionnaires. The teachers

were also together with the research contact at their school, responsible for organizing the data collections and for supporting the students under the completion of the questionnaires.

To secure the participants' privacy, a randomly derived ID-code was used as an anonymous substitute for name or e-mail address. The name of the school was made anonymous through the use of a two-digit code. The schools were responsible for keeping the coupling lists locked up in a fireproof cabinet and for destroying them after the last assessment point. Students and parents returned the completed questionnaires in anonymous and sealed envelopes before delivering them to the teachers and NUBU (by post), respectively. All student and staff questionnaires were returned to NUBU in registered packets where the envelopes were opened and scanned by trained staff.

Additionally, two separate password-secured databases were developed for the study to ensure safe handling of the data and the contact information. The databases were accessible for the research coordinator only. The researchers had no contact with any of the participants, and they only had access to anonymous data. The intervention model adaptors, developers and purveyors at NUBU were neither involved in the design and conduct of the study nor in the data-analyses and publication of the research findings.

5. Results

This chapter summarizes the aims, measures, and main findings of the five published papers forming the basis of this dissertation.

Paper 1 gives answers to research question 1, ˙How can we produce valid effect estimates when a preferred randomized controlled design fails'. Paper 2 gives answers to research question 2 regarding the ˙Three-year impacts of N-PALS on student problem behavior and classroom climate' while Paper 3 concentrates on research question 3 regarding the ˙Three-year impacts of N-PALS on teacher efficacy and behavior management practices'. Paper 4 gives answers to research question 4 regarding the 'Immediate impacts of PPBS on student problem behavior, classroom climate, teacher efficacy and behavior management practices'. The 5th paper concentrates on research question 5 related to ˙Addressing an underlying assumption in SWPBS/N-PALS'.

5.1 Paper 1 - Reducing Threats to Validity by Design in a Nonrandomized Experiment of a School-Wide Prevention Model

5.1.1. Aims. The main purposes of Paper 1 were to describe the strengthened nonrandomized experimental design (or nonequivalent comparison design, NEC) with active controls utilized in the longitudinal effectiveness study of the N-PALS model, and to address the likelihood of producing valid effect estimates. Whether our efforts to strengthen the quasi-experimental study design contributed to reduce plausible threats to the internal validity stemming from selection bias was tested by comparing the intervention and control schools prior to the start-up of the intervention. With reference to the generalizability (external validity) of the study results, statistical power was calculated upfront, and the representativeness of the participating

schools was examined in relation to the average of Norwegian schools using national register data. This is the first paper describing a school-based evaluation study based on a quasi-experimental design in which plausible threats to the validity were deliberately attempted minimized or eliminated in the design phase of the study. In the present study, only a few of the schools initially invited to participate in a randomized controlled trial (RCT) were willing to participate due to the research design. Alternative designs were considered, and a quasi-experimental design was seen as the most feasible. Quasi-experiments and RCTs are equal, except for the random allocation in RCTs. However, in nonrandomized studies the outcomes often can be disputed due to several potential competing explanations related to that threats to the internal and external validity have been ignored. Internal validity refers to the extent to which the results of a study can be reliably attributed to the intervention under evaluation, i.e., whether the observed relation between A (the intervention) and B (the outcome) reflects a causal relationship from A to B. External validity concerns the extent to which results can be generalized beyond the given study context (Shadish, Cook, & Campbell, 2002).

5.1.2 Procedures to strengthen the validity of the study. The current paper describes how several design elements were added to the design of our study in order to reduce or eliminate potential threats to the validity of the study outcomes (more details, Chapter 4). In addition to selection bias (occurs when the groups to be compared are not probabilistically similar to each other), other threats to the validity are: history (i.e., events independently occurring during the study period that can affect the participants ratings and thereby the study results), maturation (i.e., naturally occurring changes in participants physical, intellectual or emotional characteristics), testing (i.e., a pretest might increase or decrease a partisipant's sensitivity or responsiveness to the experimental variable), instrumentation (i.e., changes in instrumentation across assessments that may produce illusory changes in outcomes), regression (i.e., tendency

of extreme scores to approach the mean score), attrition (i.e., loss of participants to intervention or assessment is a threat to validity if the dropouts are significantly different from those who remain), and contamination (i.e., significant elements of the intervention are adopted by the comparison group during the study period).

In line with methodological recommendations (e.g., Shadish, 2011, Shadish et al., 2002), efforts were made to a) secure sufficient statistical power to detect intervention effects in the small to moderate range, and design elements were added up front, such as b) random invitation of schools following c) predefined inclusion and exclusion criteria, and d) a stratification and matching procedure, e) use of three instead of two cluster groups, f) use of both internal and external controls, g) active rather than passive controls, h) use of relevant and reliable measures, i) multiple assessment points, and j) multiple informant groups. On this basis, the strengthened longitudinal study design included two school groups respectively implementing the full three-level N-PALS model (n = 28) or an abbreviated version of N-PALS called PPBS (n = 17, referred to as BSBS in Paper 1), in addition to a control group continuing their ‘practice-as-usual’ (n = 20, referred to as BAU,‘ business-as-usual’ in Paper 1). Based on an applied open cohort design, data were collected from all teachers, assistants, principals, after-school personnel, and from all students with parental consent in grades 4-7 on totally six assessments points over five school years.

Statistical power and schools as the unit of analysis. A pertinent problem in studies where schools are used as the unit of analysis is to estimate how many clusters (schools) should be included in the study to obtain sufficient statistical power to reliably analyze program effects. Low statistical power is a major threat both to the statistical conclusions regarding effects and the generalizability of the outcomes. In the present study, the intervention is implemented at the school level (schoolwide), so the relevant allocating units were schools rather than individuals. Clustering impacts the statistical power, and the

intraclass correlation coefficient (ICC) and the Design Effect were used to adjust for clustering (Ukoumunne et al., 1999). The ICC was defined as the portion of the true total variation in outcome that could be attributed to differences between the clusters. With a calculated average school size of 245 (according to the Norwegian standard), the adjusted sample size estimation indicated that 64–76 schools were needed to achieve sufficient statistical power.

Minimizing selection bias. The recruitment process was carried out in eight steps. First, a stratifying procedure, using school size and geographical location as stratifying variables, was implemented to increase the probability that the comparison groups would match the intervention group on key characteristics. Second, a sample of elementary schools (grades 1–7) in 17 strategically selected municipalities were carefully informed and invited as N-PALS or control schools (PAU). School size was used as a stratification variable because it was expected that size would be related to initial school variation in behavior problems, main outcomes, and implementation quality (fidelity). Prior to invitation, all elementary schools (N = 165) in the 17 municipalities were therefore divided into three groups based on school size. Third, schools that, according to their home page (validated by phone or e-mail), were actively implementing other structured school-wide or community-wide programs were excluded in order to avoid program contamination. Fourth, very small schools with fewer than 100 students were excluded. This resulted in a total of 126 schools considered eligible for the study. Fifth, each of the eligible schools was assigned a number and randomly invited by blind drawing of numbers from a container. Approximately 50% oversampling was used. Sixth, another subsample of 44 schools located in the same municipalities as the intervention schools were randomly selected and invited as controls (PAU). Prior to the invitation, the schools in the control condition were matched to the N-PALS group on school size and on geographical location. To minimize the possibility of program contamination, only schools

located more than 5 km from the nearest N-PALS school were invited. Seventh, the schools in the N-PALS short version (PPBS) were recruited the following year by an open invitation to all eligible elementary schools in four selected urban municipalities with no N-PALS or PAU schools, more than 100 students, and no ongoing implementation of school-wide programs. The invitation was advertised in relevant journals and on the NCBD's homepage, a flyer was distributed, and the local educational authorities were informed by e-mail. In all, 17 schools were recruited in this group, but no matching to the N-PALS group occurred.

Multiple comparison groups. Rather than a single comparison group, using multiple comparison groups that are deliberately selected to overlap with the intervention group can expand the researcher's ability to explore more threats to the causal inference and to triangulate toward a narrower bracket within which the effects are inferred to lie. That is, controls should come from the same location as the intervention group and should share as many observed substantive characteristics as possible at baseline (Shadish, 2011). The current study also used a three-group comparison design, and by including both an alternative intervention and a control group, the probability of bias stemming from the "Hawthorn" or "Novelty" effects were reduced. Likewise, to yield more accurate results and realistic effect estimates, we decided to operate with both a local and an external comparison group, which also served as active rather than passive comparisons.

Multiple informant approach. No informant group was expected to provide more valid and reliable data than others (Kraemer et al., 2003) so the study called for a multi-informant approach. Our approach involves data from several informant groups including the principals, teachers, assistants, after-school personnel, parents, and students. Based on previous research, only moderate interrater agreement should be expected when multiple informants contribute with information (Kraemer et al., 2003). Discrepancies in behavior ratings across informants may stem from influences of the actual students, the contexts or situations in which the

students are observed, the perspectives and experience of the different raters, and from measurement error. The use of multiple informant groups was expected to reduce bias stemming from testing and remove the risk of mono-informant bias.

Instrumentation. The study variables were selected to match the expected program outcomes at the school, group, and student level, and to allow examination of factors and circumstances that might moderate, mediate, or confound the outcomes. The main focus was on measuring the level and dimensions of student problem behavior in schools across time as well as individual development of behavior problems, academic performance, and social skills. Emphasis was placed on the use of well-established key measures with sound psychometric properties. When adequate measures were not available, new measures were developed and piloted prior to use.

Multiple assessment points. In order to reduce threats to internal validity stemming from maturational trends, regression artifacts, and study testing as well as attrition related to intervention a double pretest, two intermediate measurement points, were included in the design, resulting in a total of six measurement points (Shadish et al., 2002; Ukoumunne et al., 1999). To prevent biased dropout and promote high response rates, several strategies were added (Wineman & Durand, 1992) like appointing a research contact at each school, the respondents were free to choose between paper and digital format, written, standardized instructions were provided to all participants and all information for parents were translated to the four most frequently spoken foreign languages: English, Urdu, Somali, and Bosnian.

5.1.3. Outcomes. To test the representativeness of the study sample and if the three cluster groups (N-PALS, PPBS, PAU) initially differed on outcome variables and on significant municipality, school, student, and staff characteristics, a series of variance, chi square and multi-level analyses were run on baseline data collected from the principals, school staff, class

head teachers and students in the 65 participating schools. In addition, data from national registers were used. The assumption was that few and small differences between the three school groups indicated that the strengthened design reduced the likelihood of selection bias and increased the likelihood of valid effect estimates.

Comparing more than 50 school, staff, student, and learning environment variables, significant group differences were found on only four, which could be expected by chance due to the many comparisons. No group differences were found on any of eight potential primary outcome variables related to student problem behavior, while slightly lower baseline scores were observed in the N-PALS group than in the two other groups on three of 11 potential secondary outcome variables (classroom climate, teacher collective efficacy, self-efficacy). Additionally, the initial analyses based on national register data indicated acceptable external validity in that the 65 participating schools and the 21 communities in which they were located only differed significantly on one of the 20 variables tested. Moreover, the baseline comparisons supported that school size was a relevant stratifying variable. Results from variance analyses showed, as expected, several significant school-size differences, and results from an additional hierarchical regression analysis showed that the school problem pressure[4] as reported by the principals (11 items, $\alpha = .77$) increased as a function of school size, low perceived teacher collective efficacy, low parent education, and high population density in the municipality.

5.1.4. Discussion. Taken together, the baseline analyses indicated that the three cluster groups

[4] School problem pressure was measured as a function of the 1) proportion of special education students, 2) proportion of students referred to the school educational services, 3) proportion of students referred to the child welfare or mental health services, 4) proportion of students reported to the police, 5) proportion of students with sever academic problems, 6) proportion of students with severe and 7) moderate behavior problems, 8) proportion of students with weak social skills, 9) proportion of students transferred to another school or class due to behavior problems, 10) proportion of students with so severe behavior problems that they should not have been in an ordinary school, and 10) proportion of students who are fully or partly trained outside their ordinary class.

were more similar than different, in that they seemed similar on approximately 90% of the variables tested. The observed findings point to that the efforts to strengthen the nonrandomized cluster group experiment increased the comparability of the groups, and accordingly, the internal validity or trustworthiness of the outcomes. The baseline comparisons also indicated the study sample to be representative, pointing to an acceptable level of external validity (at least in the Norwegian context). Although several threats to the validity likely have been reduced, we cannot claim that they were fully eliminated. Neither can we rule out the chance of group-differences on unmeasured variables related to the effect estimates.

5.2 Paper 2 - School-Wide Positive Behavior Support – Norway: Impacts on Problem Behavior and Classroom Climate

5.2.1 Aims. The primary purpose of the present study was to examine the effectiveness of the three-tired N-PALS model on student problem behavior and on the learning climate in class. A second aim was to investigate whether the quality of implementation had moderating effects on the outcomes. The research questions were: a) do schools that implement the three-level N-PALS model demonstrate lower prevalence of teacher/staff rated student problem behavior inside and outside the classroom over time and a more positive learning climate in class than comparison schools?, and b) are the intervention outcomes moderated by implementation quality?

5.2.2. Method. A strengthened non-randomized experimental design was used in the study. Several elements were added to the design to reduce potential threats to the internal validity. Measures were taken to secure sufficient statistical power to detect intervention effects in the

small to moderate range, and schools were randomly invited according to predefined inclusion and exclusion criteria. Moreover, a stratification and matching procedure was applied, the use of relevant and reliable measures, and multiple informant groups at multiple measure points. Additionally, an open cohort design was applied, in which that new staff members and new 4th graders could be enrolled at each measure point. Questionnaire data were collected from staff members (principals, teachers, assistants, special education teachers, after school personnel) and students in grades four to seven (9-12 years) at four measurement points each (T1-T4) across four successive school years in the 28 participating N-PALS schools and the 20 comparison schools doing "regular practice". The study had a double baseline with T1 (staff ratings only) and T2 (T1 for students) six months later, at the beginning of a new school year and close to the initiation of the intervention. T3 (T2 for students) was conducted after one year of implementation while post-test (T4) was conducted after three years of implementation. The students also contributed at a second intermediate measure point after two years of implementation. The outcome evaluation was conducted after one and three years of experience with N-PALS.

School and student characteristics. Participants were 48 Norwegian primary schools (grades 1-7) with a total number of 13,570 students in 1st to 7th grade at baseline (N-PALS = 7,964, control = 5,606), of whom 51% were boys and about 6% had ethnic minority background. Twelve of the schools were considered small (less than 200 students), while 24 were of medium size (201-350 students) and 12 (351-780 students) of large size. No initial group differences were detected between the intervention and comparison schools, except for in reading performance (test scores in 5th grade) where the N-PALS group scored lower than the comparison group, $F(1,47) = 4.81$, $p = .03$). The participating schools did not differ from the national average on standardized test performance scores nor on student ratings of the learning environment from the annually conducted National Student Survey (Statistics

Norway, 2009).

Staff characteristics. At T1 the total staff amounted to 1,064 in the intervention schools and 750 in the comparison schools. About 56% of the staff had teacher training, 12% had additional special education training (minimum one year) while 4% had no formal training. No group differences were found in the staff characteristics at baseline. The actual study sample counted 1,266 persons at baseline, of which 1,211 (96%) contributed to the study. In total, significant group differences were found on only one of 29 study variables. Even if potential differences on unobserved variables cannot be ruled out due to the non-randomized design, the analyses indicated that the groups were initially comparable on the variables assessed in this study and allowed for meaningful between-group comparisons of change across time.

Intervention. The core components of N-PALS as described in a handbook are: 1) school-wide positive behavior support strategies including teaching of school rules, positive expectations and social skills, and systematic praise and encouragement of positive behavior (including reward cards), 2) monitoring of student behavior on all arenas of school by using the School-Wide-Information system (SWIS), 3) collectively applied school-wide corrections with mild and immediate consequences (response cost), 4) time-limited small group instruction or training in academic or social topics, 5) individual interventions and support plans, 6) classroom management skills for teachers, and 7) parent information and collaboration strategies. Generally, it is expected, that it takes three to five years to fully implement the SWPBS/N-PALS model.

Measures. The prevalence of student problem behavior was measured by staff ratings using ˙Problem Behavior in the School Environment last Week' (15 items, α ranged from .81 to .82) and ˙Problem Behavior in the Classroom last Week' (20 items, α ranged from .86 to .88) (Grey & Sime, 1989; Ogden, 1998). Staff members reported how many times they had

observed negative behavior incidences inside and outside classrooms during the week prior to assessment. Factor analyses revealed that both measures had two subfactors: ˙Moderate problem behaviors' and ˙Serious problem behaviors' (α ranged from .70 to .90). The scales have shown satisfactory psychometric properties in several prior Norwegian studies (e.g. Kjøbli & Sørlie, 2008; Lindberg & Ogden, 2000; Ogden, 1998; Sørlie & Ogden, 2007). A translated version of the ˙Classroom Environment Scale' (CES; Moos & Trickett, 1974) was used by teachers to assess the quality of the learning climate in class (α = .83 to .85). An equivalent 22 item student version developed by Sørlie and Nordahl (1998), was used to assess the students' perceptions of the classroom climate (*a* = .86 to .88). The scale has shown satisfactory psychometric properties in prior Norwegian studies (e.g. Sørlie & Ogden, 2007; Sørlie & Nordahl, 1998). Additionally, we asked the principals at baseline and post-test how many students that were fully or partly educated outside ordinary classroom due to problem behavior.

Implementation quality was assessed with ˙The Effective Behavior Support Self-assessment Survey' (EBS-SAS, 46 items) which was completed by all teachers and other school staff (Sugai, Horner, & Todd, 2009). EBS-SAS is a web-based questionnaire that is routinely completed once a year in all intervention schools. EBS-SAS has been used in several prior evaluation studies (e.g., Bradshaw, Mitchell, & Leaf, 2010), and measures the perceived extent to which the intervention model has been implemented with fidelity at the school level, in the classroom context, and in individual cases as well as in common areas like hallways and the playground. In the present study the alphas for the sumscale and subscales ranged from *α* = .90 to .92. In the moderation analysis only the sumscore was used. For N-PALS/SWPBS to be adequately implemented, an 80% threshold score is set as minimum.

5.2.3 Analyses. To examine differences in change across time between the control and

intervention group, longitudinal multilevel analyses with three levels were run. The data structure with repeated measurements nested within respondents (staff members or students) and groups (schools) suggested a multi-level approach. In all main effect models, the control group was set as the reference group while baseline was set as the reference time-point, and T2-T4 were included in the analyses. In addition, the potential influence of contextual variables was statistically accounted for by including school size, proportion of staff members without formal training, portion of staff working in the after school services, proportion of special education students, and proportion of students with immigrant background as covariates at the school level. All covariates were centered to minimalize multicollinearity (Graham, 2003), and non-normally distributed variables were log-transformed (i.e., severe behavior problems in school and classroom setting, behavioral correction). Because the time between measurements varied, an unstructured residual covariance structure was chosen for the level 1-residuals. At the school level, random intercepts were estimated using a scaled identity covariance structure. Main effects were investigated by adding a Time x Group interaction (control group = 0, intervention group = 1) to the models. A second series of models examined whether there were any moderating effects of implementation quality on the observed outcomes in the N-PALS group. In addition to the original variables, these models were extended by the two-way interaction term Time x Implementation Quality. All relevant lower-order terms were included to ensure balanced regressions equations. Missing data were estimated using the direct-likelihood method, assuming a missing at random mechanism.

5.2.4 Outcomes. Using an open cohort design, the attrition was as expected rather large due to normal fluctuations in the staff and student population across time. The comparison analyses revealed few differences between those participating and those missing and no systematic differences in attrition between the intervention and control group. In the

multilevel analyses of intervention effects, missing data thus were estimated using the direct-likelihood method, assuming a missing at random mechanism (MAR; Beunckens, Molenberghs, & Kenward, 2005).

Indications of positive three-year main effects of the N-PALS model were found for moderate and less severe behavior problems on common school areas and in the quality of the learning climate in class as rated by teachers. The effect sizes were relatively modest using Cohen's *d* and ranged between .17 and .25 (but large according to more recent effect benchmarks, Kraft, 2018; Lipsey et al., 2012). The multi-level analyses revealed that reduced level of problem behavior occurring on the school's common areas over time was more evident in the intervention group as compared to the control group (diff = -1.51, $p = .001$). This finding indicates a significant main school-level effect of N-PALS, and was true both for serious behavior problems and moderate behavior problems. The prevalence of problem behavior occurring within the classroom context was also substantially reduced in both groups during the study period, but no significant intervention effect was found.

The classroom climate in both school groups showed a positive trend as rated by the school staff, but a significantly more positive developmental trend was reported by school staff in the intervention condition than by staff in the control condition, indicating a positive main effect of N-PALS. On the other hand, no main effect of the intervention was observed in the students' ratings of the psycho-social learning conditions in class. However, the number of students that were pulled out of class and teached in segregated settings due to problem behavior seemed to have declined with 37.5% from baseline to post-test in the N-PALS schools, while the number increased with 54% in the control schools.

The implementation quality (fidelity) measure indicated that after three years, 75% of the intervention schools (N = 18) had implemented N-PALS with required fidelity (minimum 80% on EBS-SAS). Generally and as hypothesized, implementation quality and school size

were inversely related (r = -.25, p <.01), indicating higher implementation fidelity in smaller sized schools. Implementation quality significantly moderated the effects of N-PALS on both the primary outcome problem behavior variables, and this was true both for moderate and severe problem behavior. Implementation quality also moderated the effects on the learning conditions in class as rated by the school staff, and on the 4th-7th graders' ratings of the relationship between students in class. Additionally, the analyses revealed that the schools with higher implementation quality had a larger reduction both of serious and moderate problem behaviors on common school areas. Moreover, the greatest increase in the quality of learning climate in class as rated by staff was found in schools with the highest fidelity scores.

5.2.5. Discussion. The intervention effects were examined after one and three years of implementation and the analyses indicated a significant positive main effect of the N-PALS model across time on the level of student problem behavior occuring on common school arenas. Relative to the control schools, lower levels of all kinds of problem behavior were observed in the N-PALS schools. This was especially true for schools with high implementation quality. Improvements of students' behavior in unstructured settings such as hallways, transitions and playground are of great interest to school personnel, because they have contextual features that tend to increase student problem behavior (Algozzine et al., 2010). The positive change found in the N-PALS schools can partly be attributed to the common set of behavioral expectations, teaching of rules and social skills, consistent consequences and other aspects of the schoolwide positive behavior support model (Lewis, Power, Kelk, & Newcomer, 2002).

As concerns serious and moderate problem behavior in the classroom context, positive intervention effects were evident only in N-PALS schools with high implementation quality. Given the strong emphasis on comprehensive school-wide positive behavior support in the N-

PALS schools, a main effect of the intervention on classroom-related problem behavior could be expected. The finding indicates that the teachers in the comparison schools had equally effective methods and procedures for preventing and managing classroom problem behavior as about half of the intervention schools. So, even if there was substantial reduction in classroom problem behavior also in the N-PALS schools with lower implementations scores, this expected development was matched by a similar reduction in the comparison schools. Moreover, N-PALS seemed to have a positive inclusive impact, in that the number of students educated outside their ordinary classrooms declined from pre to post, while the number in fact increased in the control schools.

Indications of a positive main effect was also observed on the quality of the learning climate in class as rated by the teachers. Again, implementation quality moderated the outcomes. The comparisons between the intervention and control groups showed, that if implemented with high fidelity, the N-PALS model significantly decelerated and even countered the expected reducion in the students' perceptions of the psycho-social learning climate. It should be noted that all effect estimates were robust when controlling for several other possibly influencing school level factors besides the intervention model itself, including significant characteristics of the staff and student body.

In summary, all study results pointed in the expected direction, indicating positive main or differential impacts of N-PALS on student problem behavior and the learning conditions in class. But the results also raise a question that needs elaboration; why did the control group do so well? Closer examination showed that most of the comparison schools had been implementing other school-based programs or projects during the study period. Even if implementation of other school-wide programs was an exclusion criteria, the principals reported at post-test that *during* the study period all but three of the control schools had been implementing one or two evidence-based programs targeting bullying, social skills

promotion, or the learning environment. Thus, N-PALS was evaluated using a harder test than initially intended, and these group differences thus may be considered conservative estimates of the N-PALS impact. Rather than comparing the intervention schools with regular practise, they were compared to schools implementing other evidence-based interventions. It was not possible, however, to estimate how the alternative programs may have influenced the outcomes of the comparison schools.

The significant effect sizes in the current study were in the range of *d* = .13 to .25, which compares nicely with Bradshaw et al. (2012) who reported ES-values of *d* = .08 and .17 based om teacher assessed student behavior in the USA. The magnitude of intervention effects on school problem behavior in the present study also match well with the mean *d* of .20 reported in meta-analyses of universal school programs (Durlak et al., 2011; Wilson & Lipsey, 2007). The moderating effects of implementation quality on the intervention outcomes were in line with prior research and showed that the schools with high implementation scores generally benfitted more than schools with lower scores (e.g. McIntosh, Bennett, & Price, 2011; Fixsen et al., 2005).

Limitations. First, a randomized controlled design was not used in the present study. Second, even if very few significant group differences were found at baseline, selection bias may have occurred in the process of recruiting schools to the study. There may be undetected group differences on non-observed variables. Not all of those invited accepted the invitation among the N-PALS schools nor among the comparison schools, but any selection bias may have been similar in both groups in that no differences were found between the declining and accepting group. A potential third limitation is that the staff informants were also responsible for implementing the interventions. This may have created a positive response bias in the assessments of implementation quality and student outcomes in the N-PALS schools. As regards student outcomes, any positive bias may have affected both the intervention and the

comparison group equally.

In conclusion, the findings presented in Paper 2 indicated that the N-PALS model is effective in reducing the level of more and less severe student problem behavior in Norwegian primary schools and in promoting qualitatively better learning conditions in class - on the condition of being implemented with fidelity over at least three years. The results also indicated that the three-level prevention model may increase the schools' ability to reach out to all students and reduce the number of students singled out for individual intervention plans. All effect estimates were robust when controlling for several other possibly influencing school, student, and staff body factors.

5.3 Paper 3 - Examining Teacher Outcomes of the School-Wide Positive Behavior Support Model in Norway: Perceived Efficacy and Behavior Management

5.3.1 Aims. This study investigated the potential impacts of the N-PALS model on staff perceived collective efficacy, self-efficacy, and behavior management. The research questions were as follows: a) Do staff in schools that implement the N-PALS model over a three-year period demonstrate higher perceived efficacy (self-efficacy and collective efficacy) across time than staff in control schools doing ˙practice-as-usual'?, b) Do the staff in N-PALS and control schools differ in behavior management (i.e., use of positive behavior support and behavioral corrections) as reported by students and staff?, and c) Are intervention outcomes moderated by implementation quality?

5.3.2. Method. The effectiveness of the three-level school-wide prevention model, N-PALS, was examined using longitudinal data from the same sample and informants as in Paper 2. Questionnaire data were collected from staff and students (grades 4 to 7) at four measurement

points across four successive school years in 28 intervention schools and 20 comparison schools.

Perceived ˋCollective Efficacy in Schoolˊ was measured with a 12-item translated version (Goddard, 2002) of the well-validated ˋCollective Efficacy Scaleˊ (Goddard, Hoy & Woolfolk Hoy, 2000). The measure assess the extent to which a faculty believes in its joint capability to positively influence student learning. The internal consistency scores across assessment points (α) ranged from .82 to .85. A recent study (the first conducted in Scandinavia) using staff ratings (n= 1,528) from the full sample of 65 schools in the N-PALS evaluation study, examined the structural, concurrent, and predictive validity of the short scale, and concluded that it is both a valid and reliable measure in this context (Hukkelberg & Sørlie, in review). Perceived ˋSelf-efficacy in Schoolˊ was measured by a 30-item scale developed for the study. Each staff member was requested to rate how competent (skillful) she or he felt in managing and preventing problem behavior and promoting the students' academic skills. Chronbach's alpha was .96 at all four assessment points. Behavior management was assessed by staff and students using the following scales developed for this study: ˋPositive Behavior Supportˊ (staff version: 9 items, α = .74 to .76; student version: 16 items, α = .88 to .92) and ˋBehavioral Correctionˊ (staff version: 8 items, α =.57 to .66; student version: 5 items, α = .58 to .61). We hypothesized that in order to reduce the amount of student problem behavior, the ratio between positive behavior support and behavioral correction is essential. A rule of thumb in SWPBS/N-PALS is that teachers should encourage and reward the students' positive behavior four to five times for every behavioral correction. ˋImplementation qualityˊ was assessed with the same measure as in the previous study using the EBS-SAS scale (Sugai et al., 2009).

5.3.3. Analyses. To examine differences in change across time between the control and

intervention group, longitudinal multilevel analyses (three levels) were conducted. The data structure with repeated measurements nested within respondents (level 2, staff, students) and groups (level 3, schools) suggested a multi-level approach. Because the spacing between measurements varied, an unstructured residual covariance structure was chosen for the level 1 residuals. At the school level, random intercepts were estimated using a scaled identity covariance structure. Main effects were investigated by adding a Time x Group interaction to the models. A second series of models examined whether there were any moderating effects of implementation quality in the N-PALS schools. In addition to the original variables, these models were extended by the two-way interaction term Time x Implementation Quality. All relevant lower-order terms were included to ensure balanced regressions equations. Separate analyses were run for the teacher and student rated outcome variables. Missing data were estimated using the direct-likelihood method, assuming a missing at-random mechanism.

5.3.4. Outcomes. At the first assessment, 75% of the staff members participated and 67% participated at post-test. Of the students, 94% with parental consent participated at project start, and 91.9% participated at post-test. Variations in participation among staff were primarily attributed to staff quitting their jobs (N = 560) while new staff members were successively trained in N-PALS and recruited for the study. Missing data and varying student participation rates were primarily due to the open cohort design, which implied that new 4th graders and other students new to the schools were included at each measure point whereas older students successively left the study by the end of 7th grade.

Results from the multilevel analyses showed that there was a general positive growth across time in the school staffs' collective efficacy scores. After three years of implementation, the positive change across time was significantly more evident in the intervention group than in the control group. A similar progress was observed for self-

efficacy, indicating significant main effects of N-PALS on these variables. A main intervention effect was also observed for disciplinary practices: the staff in the N-PALS schools increased the amount of proactive and supportive practices considerably more from pre to post assessment than did their colleagues in the comparison schools. Although there was a general decrease in the use of behavioral corrections across time, no significant group differences were observed. However, according to the staff ratings, the ratio between positive supports and corrections in the N-PALS group increased from 2.5 at pre-test to 4.5 at post-test compared to from 2.5 to 3.1 in the comparison group. The student ratings of the teachers' behavior management practices indicated no significant group differences in change across time. Additional analyses revealed no impacts either by gender or class level on the students' ratings. Moreover, no significant moderation effect of implementation quality was identified in any of the outcome variables. The significant intervention effects ranged from small to large according to Cohen's *d* which ranged between .14 and .91.

5.3.5 Discussion. In conclusion, the SWPBS model as implemented in Norway appeared effective in promoting school staffs' perceived individual and collective efficacy. According to the staff, the model also appeared effective in changing their behavior in a more positive direction. However, positive change in staff ratings did not translate to the students' ratings. In contrast to what was expected from prior research and recent analyses of student outcomes of N-PALS (c.f., Paper 2), implementation quality did not moderate the staff-related outcomes (i.e., all intervention schools seemed to benefit approximately equal).

Limitations. Some limitations of the study should be mentioned. First, although few significant group differences were identified at baseline, selection bias may have occurred in the process of recruiting schools to the study. A second limitation is that the staff in the N-PALS schools was responsible for implementing the interventions and felt allegiance to the

model. This may have created a positive response bias in the assessments of implementation quality and staff outcomes in the N-PALS schools. The lack of coherence between the staff- and student-based effect estimates may of some be perceived as a third limitation. A possible fourth limitation relates to EBS-SAS as the sole implementation measure as it was originally developed for needs assessment, not specifically as a fidelity measure.

5.4 Paper 4 - Preventing Problem Behavior in School Through School-Wide Staff Empowerment: Intervention Outcomes

5.4.1 Aims. The main aim was to examine if there were any school level effects of the abbreviated version of N-PALS, the ˙Preventing Problem Behavior in School' (PPBS) intervention on student problem behavior, the learning conditions in class, the school staff's perceived collective and individual efficacy, and on their disciplining practices (i.e., use of positive behavior supports and behavioral corrections) as rated by the staff members themselves and by students. A second aim was to investigate if some schools benefitted more from the intervention than others (i.e., differential effects)

5.4.2. Intervention. The PPBS intervention was based on the same principles as the full-scale N-PALS model but focused on the universal level mainly, including: 1) a school-wide approach and differentiated evidence-based practices, 2) systematic positive reinforcement of expected pro-social behavior, 3) corrections (mild consequences) following problem behavior, 4) good directions, and 5) establishing a functional support system. The PPBS included a 30-hour in-service training program for school staffs that lasted four full days and was locally organized with 1-7 schools per site. The entire school staff was included in the program training which was provided by staff from The Norwegian Center of Child Behavioral

Development, free of charge. The participants were provided with an intervention manual, and all of the training materials could be downloaded from the Internet. The standardized training sessions were led by the program developers and were composed of a combination of lectures, demonstrations, training, coaching and "homework." In addition, unlike in the full-scale version, the PPBS schools did not have access to the SWIS component or received external supervision, training in interventions relevant for students at risk, or any technical support. Neither did the PPBS schools use systematic data on discipline referrals or functional behavioral assessments to determine appropriate interventions. Nor did they have access to measures of implementation quality or Web-based information to evaluate their outcomes and progress.

5.4.3. Method. The 17 intervention schools were compared to the same control group as in the evaluation of the full three-level model (PAU) using the same measures and informant groups (see Paper 2 and Paper 3). In addition, training dosage was measured, referring to the intervention school's mean training attendance score. Dosage was calculated on each staff member's participation across the four training days and aggregated at the school level. Data for the current study were collected in three waves: T1 (staff only) was at the end of the school year prior to the implementation of PPBS, T2 was close to the initiation of the intervention while T3 (post-test) was four months after the program training.

5.4.5 Analyses. Multilevel analyses (level 1: T1, T2, T3, level 2: staff or students, level 3: schools) were conducted in order to examine differences in change over time between the intervention and control groups. At the time of the first assessment, 77% of the totally 1,266 staff members in the PPBS and control schools were participating. Overall, 1,308 staff members participated at one or more time points. The staff response rate was above 80%

across assessment points.

At baseline there were totally 6,172 students (PPBS = 3,087, control = 3,085) in grades 4-7; the parents of 14% of those students did not allow them to participate while 9.9% never returned the consent form. Evenly distributed across grades, 4,687 students contributed to the study at pre-test and 4,630 contributed at post-test. The response rates were high (95.5% and 91.5%). The attrition was considered modest and reflects the study's open-cohort design, that two control schools dropped out and normal fluctuations in the staff and student populations. Because few differences between participants and non-participants were identified, missing data were estimated using the direct-likelihood method (MAR, Beunckens et al., 2005).

5.4.6 Outcomes. Indications of a main effect of the PPBS intervention were observed in staff-reported problem behavior in common school areas (T1-T3 diff = -2.50, p = .000), which was true both for severe (T1-T3 diff = -0.28, p = .020) and moderate problem behaviors (T1-T3 diff = -2.23, p = .000). The prevalence of problem behaviors observed in the classrooms was also reduced during the intervention period (PPBS T1-T3 change = -2.98, control T1-T3 change = -1.81), and indications of a marginally significant PPBS effect were found (T1-T3 diff = -1.17, p = .058). This reduction across time was primarily related to reduced occurrence of severe problem behaviors (T1-T3 diff = -0.46, p = .037). No main effect on the classroom climate was found, either as rated by staff or students. Moreover, the multilevel analyses based on staff reports indicated a significant main effect on disciplinary practices in the schools. The staff in the PPBS schools increased the number of positive behavior supports from baseline to post-test more than did their colleagues in the control schools (T1-T3 diff = 2.90, p = .000). Although there was a general decrease in the use of behavioral corrections, no significant group difference was found. Likewise, student ratings of the teachers' behavior

management showed no significant group differences in change over time. Finally, marginally significant main effects of PPBS on staff rated collective efficacy (T1-T3 diff = 0.79, *p* = .063) and self-efficacy (T1-T3 diff = 1.84, *p* = .062) were observed.

Moreover, there were indications of differential intervention effects. Moderation analyses indicated that when compared with control schools of same size, the small-to-medium PPBS schools benefited more than larger schools in terms of reduced problem behavior on common school areas, better classroom climate, and higher perceived collective efficacy. Additionally, PPBS schools with a high mean training dosage demonstrated a significantly greater reduction in student problem behavior and a greater increase in collective efficacy than schools with a low dosage. Besides, intervention schools with high implementation quality showed greater improvements in classroom climate and positive behavior supports than did schools with low fidelity scores. Compared with control schools on varying proportions of unqualified staff, the positive changes across time in both the quality of the learning conditions in classrooms and the staff's use of positive discipline were significantly greater in PPBS schools with high proportion of staff members who lacked formal training than in schools with lower proportions.

The effect sizes in the present study ranged from *d* = .15 to .41. The four months impacts of the abbreviated version of N-PALS (PPBS) on the level of problem behavior in school appeared relatively equal to the three-year impacts of the N-PALS, whereas the full-scale model seemed to ˙outperform' the abbreviated version concerning effects on the learning conditions in classrooms and on staff outcomes (collective efficacy, self-efficacy, behavior management).

5.4.7 Discussion. Taken together, the outcome analyses indicated that PPBS had some promising, immediate effects. The moderation analyses indicated that to be optimally

effective, high attendance by a school's entire staff during training sessions and high fidelity across all school settings are required. The moderation analyses indicated that it is likely more difficult to induce school-wide changes in large schools than in smaller schools, and that more comprehensive implementation supports may be required to "turn large schools around." Moreover, the results indicated that PPBS may positively contribute to increasing school staff's competence and practical skills in the prevention and handling of student problem behavior—particularly among those staff members who do not have formal training as teachers. Based on the indicative intervention effects, we argue that the PPBS intervention may be appropriate as a basic approach to facilitate the development of school-wide or more common attitudes and strategies in the prevention of behavioral misconduct. For some schools (e.g., large schools, schools facing stable high levels of problem behavior), PPBS may function as a "springboard" to more comprehensive approaches such as the N-PALS full-scale model. However, the study results must be replicated, preferably in randomized controlled trials, before more firm conclusions regarding the PPBS's effects can be drawn. Whether the observed intervention effects can be sustained over a longer time span also should be tested in future studies.

5.5 Paper 5 - Multilevel Analysis of the Relationship Between Teacher Collective Efficacy and Problem Behavior in School

5.5.1 Aims. An underlying assumption in the N-PALS model stemming from social cognitive theory and a weighty argument for school-wide prevention of behavior problems among children and youths, is that the culture of a school as an organization has significant impacts on how the students behave – for good or bad. In Paper 5 it was for the first time empirically tested if perceived teacher collective efficacy might be considered an indicator of the school

culture (i.e., an organizational factor), and whether collective efficacy has an impact on the prevalence of student problem behavior - concurrently and over time. Collective efficacy in school is defined as the teachers shared beliefs about their combined capability to organize and execute courses of action required to produce student success (Bandura, 1997; Goddard, Hoy, & Woolfolk Hoy, 2004).

5.5.2 Method and hypotheses. Teacher collective efficacy and self-efficacy (independent variable, referred to as self-perceived teaching competence in Paper 5) were measured using the same rating scales as in Paper 3 and 4. The level of problem behavior in school was, as in Paper 2 and 4, rated using two scales developed by Grey and problem Sime (1989, i.e., problem behavior in the school environment last week and behavior in the classroom last week; dependent variables). Additional covariates were prior levels (T1) of the dependent variables, school size, proportion of special education students and of ethnic minority students, plus teacher gender, age, educational background, and years of experience. To account for the hierarchical data structure (teachers nested in schools) and decide on a variance structure that accounted both for within-school and between-school variance, data were analyzed using a combination of a variance component model and multilevel regression models (for more details, see paper).

It was expected that both teacher collective efficacy and student problem behavior are phenomena that vary greatly between schools. Building on social cognitive theory and prior research on the relationship between teacher collective efficacy and student academic achievement, we also expected collective efficacy to be a stable school-level construct significantly contributing to explain between-school variance in student misconduct. We assume that collective efficacy will enhance the teachers' beliefs that they can positively affect the behavior of all students and also influence what they choose to do as a group to

meet student problem behavior. For collective efficacy to be considered a sociocultural organizational factor, it should be high across-time stability as well as between-school differences in the teachers' ratings. We hypothesized that schools with high collective efficacy would have less student problem behavior than schools with low collective efficacy. We expected the two phenomena to be correlated when measured at a given time point (concurrent relationship). We also expected that collective efficacy would function as a predictor (protective factor) of behavior problems in school over time (temporal relationship). However, it can also be the other way around, that experiences of high levels of behavior problems in school reduce the teachers' sense of collective efficacy. In other words, it might be a reciprocal relationship between collective efficacy and student misconduct.

5.5.3. Outcomes. Building on pre-intervention data collected across two school years (T1 and T2, 6 months lag) from 1,100 teachers (data from other staff members excluded in this study) in 48[5] primary schools participating in the longitudinal N-PALS evaluation study, the study results showed that;

- there was high reliability across raters in collective efficacy,
- collective efficacy as reported by teachers is a reliable school-level indicator,
- between-school differences in collective efficacy were significant and stable across time points (at least from one school year to another), while the within-school stability seemed somewhat lower,
- student problem behavior also varied significantly, both within and between schools, but was not a reliable indicator of school-level problems,
- schools with high collective efficacy generally had lower levels of problem behavior,
- schools with high prevalence of problem behavior in the classroom context often had

[5] Anonymous data from PPBS-schools were not available to the researcher when these analyses were run.

additional high prevalence of problems on common school arenas, such as school yard and hall ways,

- teacher collective efficacy predicted both moderate and severe problem behaviors in the classrooms context as well as on common school premises,
- changes in the prevalence of school problem behavior over time systematically predicted changes in the teachers' collective efficacy perceptions.

5.5.4. Discussion. In conclusion and in line with the hypotheses, the results from the present study suggested that there is a strong, inverse, consistent, and reciprocal relationship (concurrently and across time) between collective efficacy and problem behavior in school. The results were robust to controlling for key school characteristics, such as school size, proportion of special education students, proportion of students with ethnic minority background, and to teacher characteristics, such as gender, age, educational background, years of experience, and self-efficacy. The findings also indicated that teacher collective efficacy is a significant and reliable indicator of a school's culture.

6. Discussion

The main purpose of the present project was to investigate school level impacts of N-PALS, which is an acronym for the adapted version of the SWPBS (School-Wide Positive Behavior Support) model in Norway. Intervention effects were tested both for the standard, full three-level model and an abbreviated version called the 'Preventing Problem Behavior in School' (PPBS). A strengthened quasi-experimental design and multi-level analyses were applied in order to examine group differences in change over time on significant student and staff related outcome variables. Moreover, the relevance of an underlying assumption of SWPBS was tested; that social characteristics of the school organization such as teachers' perceived efficacy, have significant impacts on the students' behavior. In this chapter, the research findings are summarized and commented on. The results are discussed and related to prior research, under headings of 1) effects on externalizing problem behaviors in school, 2) effects on classroom climate and inclusion, 3) effects on collective efficacy, self-efficacy, and behavior management, 4) teacher rated problem behavior and collective efficacy, 5) informant discrepancies, 6) comparing implementation strategies, and 7) magnitude of intervention effects. The chapter is rounded off with sections on study strengths and limitations, and implications for future research and practice.

Taken together (see Table 6.1), the study findings presented in Paper 2-4 demonstrate that Norwegian elementary schools implementing the three-level N-PALS model had better outcomes than comparable schools. Significant effects were found on eight of the eleven outcome variables measured three years after the introduction of N-PALS, and implementation quality moderated the effects on most variables. Moreover, indications of significant positive short-term effects (after four months) were found for the abbreviated or short version of N-PALS (PPBS) on five of ten outcome variables. Moderating effects of

implementation quality, school size, program training dosage, and/or portion of unqualified staff were found for most variables. No negative effects emerged for either of the interventions.

Table 6.1. Summary of Main and Differential Effects of N-PALS and PPBS

	N-PALS (3 years, p)		***PPBS (4 months, p)***	
Outcome variables	Main effects	Differential effects	Main effects	Differential effects
Problem Behavior in the School Environment	.001	.000 (imp.qual.)	.000	.017 (size)
- moderate	.001	.000 (imp.qual.)	.000	ns
- severe	.033	.000 (imp.qual.)	.020	ns
Problem Behavior in the Classroom Context	ns	.000 (imp.qual.)	ns	ns
- moderate	ns	.000 (imp.qual.)	ns	ns
- severe	ns	.000 (imp.qual.)	.037	ns
Classroom Climate (staff)	.017	.006 (imp.qual.)	ns	.004 (imp.qual.) .001 (size) .015 (unqual.)
Classroom Climate (student)	ns	.ns	ns	ns
- student-student relationships	ns	.000 (imp.qual.)	ns	ns
- student teacher relationships	ns	ns	ns	ns
Inclusion	.01	-	-	-
Teacher Collective Efficacy	.000	ns	ns	.014 (imp.qual.) .000 (size) .000 (dosage)
Teacher Self-efficacy	.014	ns	ns	
Positive Behavior Support (staff)	.000	ns	.000	.016 (unqual.) .017 (dosage)
Positive Behavior Support (student)	ns	ns	ns	ns
Behavioral Corrections (staff)	ns	ns	ns	ns
Behavioral Corrections (student)	ns	ns	ns	ns

ns= non-significant (p>.05)

6.1 Effects on Externalizing Problem Behaviors in School

The main objective of the standard N-PALS model and the short version (PPBS) is to prevent and reduce the prevalence of externalizing student problem behavior on all school areas. Compared to the control group, the staff in the intervention schools reported greater decreases

over time for the prevalence of moderate and severe problem behaviors occurring on common school areas.

For the standard N-PALS model, positive main effects were found both one and three years after implementation. Moderation analyses revealed that N-PALS schools with the highest implementation quality benefitted the most from the interventions. Indications of positive effects were also found for problem behaviors occurring within the classroom context. Contrary to the most optimistic expectations, significant changes in favor of the N-PALS schools were only found for problem behaviors in the schools with high implementation scores by year three.

For the abbreviated PPBS intervention, effects were measured four months after the program training, and outcomes showed that small to medium sized intervention schools and those with a high mean training dosage and a higher proportion of unqualified staff benefitted the most as concerns problem behavior occurring on school-wide areas. The significant reduction from pre to post assessment for problem behaviors occurring in the classroom context was restricted to severe problem behaviors.

Putting these outcomes into perspective, they indicated that the full scale model had more comprehensive effects on externalizing problem behavior than the abbreviated version. The longer term effects were particularly beneficial in schools that had implemented the model with fidelity. Even if the short version of the model also demonstrated short-term positive effects, these were limited to the most serious behavior problems. For both versions of the N-PALS model, the results were particularly positive for problem behaviors occurring on common school areas. An explanation for this may be that most elementary teachers succeed in controlling their classes during lessons, while the monitoring and handling of student problem behaviors in less structured situations, such as in hallways and outdoor areas, were probably less systematic and effective prior to the implementation of N-PALS or PPBS.

So, it seems reasonable that the pre-post differences in student problem behavior between intervention schools and comparison schools would be more evident outside the classrooms.

6.1.1 Relationships to prior research. The positive effects on the prevalence of externalizing problem behaviors in Norwegian schools are in line with findings from studies with different designs and in line with reviews, and meta-analyses on the effectiveness of the SWPBS model as implemented in the USA and other countries, (e.g., Bradshaw, Mitchell, & Leaf, 2010; Bradshaw, Waasdorp, & Leaf, 2012; Chitiyo, May, & Chitiyo, 2012; Flannery, Fenning, Kato, & McIntosh, 2014; McIntosh, Bennett, & Price, 2011; Noltemeyer, Palmer, James, & Wiechman, 2018). However, it should be noted that most of these studies focused on the one-year impacts of the universal level (Tier 1) only. The current study is therefore one of few which has investigated longitudinal effects from 4th through 7th grade of the combined three-level SWPBS model (Humphrey, Lendrum, Barlow, Wigelsworth, & Squire, 2013).

Moreover, indications of reduced school prevalence of externalizing problem behaviors following this school-wide approach match well with findings from several meta-analyses, research reviews, and a mega-analysis of similar school-based interventions (e.g., Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011; Matjasko et al., 2012; Sklad, Diekstra, De Ritter, Ben, & Gravensteijn, 2012; Weare, 2015; Weare & Nind, 2011; Wilson & Lipsey, 2007). A more critical comment on the effectiveness was published by Chitiyo et al. (2012) prior to the current study. They reviewed 34 articles, and found that only two of the reviewed studies showed satisfactory fidelity. The authors therefore concluded that SWPBS was a rather promising approach, but that more research with more advanced methods were needed, particularly studies with rigorous research design, operational definitions of participants and practices, and use of valid and reliable measures of student outcomes. A more recent review by Lee and Gage (2020) examined 29 studies, including seven RCTs and 22

quasi-experimental group design studies. Overall, they found statistically significant reductions in school discipline and increased academic achievement with effect sizes ranging from small to medium.

We conclude by stating that the empirical data from intervention studies, meta-analyses, and research reviews support the positive behavioral findings from the present study, but they also confirm that not all outcomes are significantly favoring the intervention schools compared to other schools. Prior and parallel research also emphasize that high implementation quality generally is associated with better behavioral outcomes, and that high implementation quality is a major challenge in most school effectiveness studies, including in those evaluating the SWPBS model, such as the present one.

6.2 Effects on Classroom Climate and Inclusion

Classroom climate was measured with a student version (Sørlie & Nordahl, 1998) and a teacher version of the Classroom Environment Scale (Moos & Trickett, 1974). In line with other studies, the underlying factors were found to be "student relations" and "student – teacher relations" (Toren & Seginer, 2015). Across time, staff ratings showed that the quality of the classroom climate increased in both school groups. However, the positive trend was stronger in the intervention schools than in the comparison schools, and among these, schools with high implementation quality scores benefited the most. Student ratings revealed a positive effect of N-PALS on the relationships between classmates, but only in schools with high implementation quality, and no significant group differences were found in the student assessments of their relations to their teachers. Generally, student ratings of the classroom climate showed a downward trend across time and class levels in both groups. This finding was also reported in the N-PALS pilot study (Sørlie & Ogden, 2007) and in another study in which the classroom climate deteriorated over time from 5th to 7th grade (Ingemarson,

Rosendahl, Bodin, & Birgegård, 2019). Reports from the school principals showed that the N-PALS model had a positive impact on inclusion, evidenced by fewer students being educated outside their ordinary classroom at pre-test than at post-test. The opposite tendency was reported from the control schools.

In conclusion, the findings indicated that schools implementing the N-PALS model achieved a more inclusive culture and better classroom climate over time, and in general, the positive changes of the classroom environment in the N-PALS schools were more evident in the staff ratings than in the students' ratings. For the PPBS intervention, no main effects on classroom climate were observed after four months, neither in the staff ratings, nor in the student ratings. An exception was small and middle sized schools implementing PPBS with high fidelity, and schools with a high proportion of untrained staff, in which a significantly more positive classroom climate was reported. The lack of a main effect could mean that direct training of the whole staff for a shorter period of time was not generally effective when intervention and control schools were compared.

6.2.1 Relationships to prior research. There is growing evidence on the SWPBS models' potential to alter the overall school climate in elementary and middle schools (Noltemeyer et al., 2018). Evidence on the SWPBS model‘s effects on inclusion and on the students‘ and teachers’ perceived classroom climate is, however, scarce. Research has shown that the quality of the classroom environment is linked to children's current and future social-emotional functioning (both social skills and problem behavior as rated by teachers, parents, students themselves or by independent observers) (Brophy-Herb, Lee, Nievar, & Srollak, 2007; Evans, Harvey, Buckley, & Yan, 2009; Mashburn et al., 2008; Shin & Ryan, 2017; Sørlie, Hagen, & Nordahl, 2020). For example, increase in the teacher-student relationship quality has been related to decreases in externalizing problems such as bullying and

aggression, and to decreases in internalizing problems such as anxiety and depressive symptoms (e.g., Berry & Connor, 2010), while conflictual and close relationships between students and teachers have been linked to disciplinary problems (Hamre & Pianta, 2001) and decreases in some children's social skills (Sørlie et al., 2020). Research also has shown that an overall positive climate is related to enhanced academic achievement in math and reading (e.g., Pianta, Belsky, Vandergridt, House, & Morrison, 2008; Rucinski, Brown, & Dower, 2018). A positive classroom climate is typically described as welcoming, respectful and safe, and characterized by high levels of warmth, positive affect, teacher sensitivity and responsiveness, and by child-centered instruction. A negative classroom climate can be described as conflictual, chaotic, hostile and out of control, or as characterized by interactions colored by frequent incidences of anger, irritability, and sarcasm.

Prior to the current study, results from analyses on inclusion and classroom climate in evaluations of the SWPBS model had not been published, except for in a Norwegian pilot study indicating a marginally significant effect based on teacher reports after two years of implementation, while no effect based on student reports (Sørlie & Ogden, 2007). Evaluations of several less comprehensive programs indicate that school-based interventions have the potential to positively alter the relational and educational features of the classroom setting in elementary schools. For example, the Fast Track social competence intervention (universal component) implemented in 1st grade, produced significant positive short-term effects on four aspects of the class milieu: expressing feelings appropriately, following rules, staying focused and on task, and level of interest and enthusiasm (CPPRG, 1999). Other examples are, the Child Development Project (Solomon, Watson, Battistich, Schaps, & Delucchi, 1996), the Responsive Classroom intervention (Brock, Nishida, Chiong, Grimm, & Rimm-Kaufman, 2008), and the Chicago School Readiness Project (Raver et al., 2008) which all have been found to positively affect students' perceptions of the classroom environment.

With a few exceptions (e.g., CPPRG, 1999), including the current study, prior studies on classroom climate are based on comparisons of students' individual reports, without taking into account the non-independence in the data due to the clustering of students within classrooms (Brock et al., 2008; Brown, Jones, LaRusso, & Aber, 2010). Moreover, the current study is one of very few studies testing intervention impacts on the quality of the classroom climate after controlling for key school, teacher, and student background factors, and examining whether such factors moderate the intervention effects (Brown et al., 2010).

6.3 Effects on Collective Efficacy, Self-Efficacy, and Behavior Management

In the evaluation of the N-PALS model, perceived collective efficacy in the intervention schools showed a more positive development than in the control schools. No similar main effect was found in the evaluation of the PPBS intervention. However, some differential effects existed, in that the staff in small to moderately sized PPBS-schools, schools with high implementation fidelity, and schools with high program-training dosage changed their sense of collective efficacy more than did the staff in large schools, and in schools with fidelity scores and program-training dosage below the mean.

Moreover, a positive longitudinal main effect of N-PALS on the school staff's perceived individual efficacy was demonstrated along with a positive effect on the staff's behavior management (or disciplinary practices). A similar group difference in change were not, however, reflected in the students' assessments of the teachers' disciplinary practices. Also for the PPBS intervention, a positive short-term effect on disciplinary practices was found, while no effect on self-efficacy occurred. Common to both intervention frameworks (as compared to the control group) was an increased use of proactive and supportive teaching strategies. Although, the use of correctional strategies did not seem significantly affected by either of the approaches, the proposed ideal ratio between positive supports and corrections

(4-5:1) was achieved in the N-PALS schools.

In summary, the effect analyses revealed that N-PALS had a positive longitudinal main effect on the school staffs perceptions both of their conjoint and individual capability to positively affect the students' behavior and learning as well as on their abilities to do so. In other words, after three years with N-PALS both the individual school staff member and the collegium as a whole felt more competent and confident that they were able to provide good teaching, and also gave more positive support to all students, independent of the students' individual needs and conditions. The abbreviated PPBS intervention seemed less able to produce positive teacher outcomes than the full three-level N-PALS model.

6.3.1 Relationships to prior research. Somewhat surprising, only a couple small-scale studies have investigated if there may be favorable impacts of SWPBS on teachers' use of disciplinary practices. In a single teacher SWPBS study, the researchers uphold that more active supervision and pre-corrections from the teacher, which are two proactive strategies recommended in SWPBS, seemed related to some decrease in the student's problem behavior (DePry & Sugai, 2002). In a second study, involving four Grades 5-8 teachers in one SWPBS school, the analyses pointed to that as the teachers' use of praise statements increased, a decrease in off-task and disruptive classroom behavior was observed (Myers, Simonsen, & Sugai, 2011). The findings from these studies support the current findings of a positive main effect of SWPBS on the teachers' behavior.

The evidence of a positive effect of SWPBS on teachers' self-efficacy beliefs is somewhat more solid. Ross and Horner (2007) reported a study involving 20 teachers from four middle schools implementing SWPBS, and showed that the teachers in the two schools with high fidelity reported significantly higher perceived efficacy scores than the teachers in the two schools with low fidelity. In a study conducted by Kelm and McIntosh (2012) in two

schools implementing SWPBS with fidelity for at least five years and three comparable non-SWPBS schools in Canada, multi-level analyses revealed a significant and large group difference in perceived self-efficacy favoring the intervention schools.

Other studies have shown mixed effects when examining the effects on teacher self-efficacy of extended versions of SWPBS. In their study of a version called the 'Comprehensive, Integrated, Three-Tired Model of Prevention' in two middle schools for one year, Oakes, Lane, Jenkins, and Broker (2013) found no indications of change in the teachers' perceived efficacy. On the other hand, more positive outcomes were reported in a study with a randomized design, in which the effects of a version called PBISplus and standard SWPBS were compared (Bradshaw, Pas, Goldweber, Rosenberg, & Leaf, 2012). The researchers reported significantly greater change in perceived teacher self-efficacy in the 20 schools implementing PBISplus than in the 22 schools implementing standard SWPBS. Taken together, most American and Canadian findings point in the same positive direction as the Norwegian findings, that the SWPBS model may positively affect teachers' sense of individual efficacy.

When it comes to 'collective efficacy', the current study is the only one that has examined effects of SWPBS. Relatedly, similar tendencies have been found in studies of 'student and faculty understanding of expectations, policies, and procedures about student conduct' (Houchens et al., 2017) and organizational health (Bradshaw, Koth, Bevans, Ialaongo, & Leaf, 2008). These studies support our finding that N-PALS may have a positive influence on the schools staff's collective efficacy perceptions, but more research are needed before a causal relationship can be said to be established.

6.4 Teacher Rated Problem Behavior and Collective Efficacy

Findings from the current study indicate that perceived teacher collective efficacy impacts

the students' school behavior. The findings support the underlying assumption of the N-PALS model that the students' behavior are primarily affected by the *social* aspects of the school context. Notably, a strong, inverse, consistent, and reciprocal relationship, both concurrently and across time, between perceived teacher collective efficacy and the prevalence of more and less severe problem behaviors was found. The results indicate that in schools with high collective efficacy or positive changes in collective efficacy over time, the teachers report decreasing rates of student problem behavior. Conversely, in schools with low collective efficacy or negative change in the collective efficacy scores over time, the teachers reported a high or increasing level of problem behaviors. However, a reverse connection also was found; that increases in school prevalence of problem behavior over time were related to decreases in the collective efficacy beliefs. The results were robust to controlling for key school characteristics such as school size, proportion of special education students, proportion of students with ethnic minority background, and to teacher characteristics such as gender, age, educational background, years of experience, and self-efficacy. Moreover, results from the mixed model analyses indicated that collective efficacy is a significant and reliable indicator of a school's operative culture. The study findings thus support the hypothesis put forward by Goddard and Goddard (2001) that perceived collective efficacy is a potent characteristic of a school's culture.

6.4.1 Correspondence with prior research. The findings are in line with prior research showing that collective efficacy beliefs represent an important predictor of between-school differences in student-level academic achievement, even when controlling for e.g., SES, prior achievement, and teacher self-efficacy (e.g., Goddard, 2002). The results also fall in line with more recent research showing that a cluster of socio-cultural characteristics of the school organization (e.g., classroom climate, student-teacher relationships, behavioral expectations,

social rules, teacher self-efficacy) are important predictors of between-school variation in student misconduct (e.g., Astor, Guerra, & Van Acker, 2010; Payne, 2008; Reaves, McMahon, Duffy, & Ruiz, 2018; Steffgen, Recchia, & Viechtbauer, 2013; Thapa, Cohen, Guffey, & Higgins-D'Alessandro, 2013; Wissink et al., 2014). Most of previous research has addressed the association between collective efficacy and academic achievement. Our findings extend prior research by establishing a relationship between *externalizing problem behavior* and perceived collective efficacy in school. To our knowledge, this is the first study of such a relationship.

By focusing on how teachers' collective efficacy beliefs in addition to perceived individual efficacy influence student behavior, the importance of having a school-wide perspective becomes clearer. When expectations and rules are communicated to the students, not as an expression of idiosyncratic values and ideas, but rather as an expression of collective school values, the impact seems to increase. There is, however, a reciprocal relationship between collective efficacy and student behavior, which indicates that implementing school-wide measures like the teaching of expectations and rules may promote both positive student behavior and the collective efficacy of the teachers. Therefore, perceived collective efficacy stands out as a highly relevant outcome variable in evaluations of the SWPBS model as well as in other evaluations of school-wide intervention efforts.

6.5 Informant Discrepancies

As noted, discrepancies were observed in the staff and student-ratings of the classroom climate and the teachers' behavior management practices. This was evident both in the evaluation of the standard N-PALS model and the shorter PPBS intervention. This discrepancy is commonly demonstrated in studies with multi-informant assessments (e.g., teachers, students, parents, trained observers) (e.g., Achenbach, 2011; De Los Reyes, 2011;

Dirks, Boyle, & Georgiades, 2011; Gresham et al., 2018; Hartley, Zakriski, & Wright, 2011; Kopershoek, Harms, de Boer, van Kuijk, & Doolaard, 2016; Rescorla et al., 2013; Romano, Weegar, Babchishin, & Saini, 2018; van der Scheer, Bijlsma, & Glas, 2019; Wagner, Göllner, Werth, Schmitz, & Voss, 2016).

As demonstrated in the present study, teachers generally report classroom environments to be more positive than do students (e.g., Fisher & Fraser, 1981; Fraser & O`Brien, 1985). Besides, in a meta-analysis, Kopershoek et al. (2016) concluded that students generally report less improvement after following intervention programs than do teachers and observers. The mean overall effect sizes (Cohen's *d*) for students, teachers, and observers were respectively 0.16, 0.24 and 0.30, and 0.10, 0.26 and 0.26 for behavioral outcomes (Kopershoek et al., 2016).

Generally, teacher ratings are found to be highly reliable and valid, and often predictive of students future functioning (e.g., Desimone, Smith, & Frisvold, 2009; Flake & Petway, 2019; Gresham, Elliott, Cook, Vance, & Keller, 2010; Ogden, 2003; Rimfeld et al., 2019; Sørlie, Hagen, & Ogden, 2008). This does not necessarily mean that the student ratings are 'wrong'. The students have another position and perspective, and in addition to having different expectations, values, and evaluation standards than the teachers, they may also disagree among themselves. Their perceptions and ratings of the classroom climate or teachers' behavior management skills, may be based on experiences across several lessons, teachers, and classmates.

There may also be challenges to the measurement process (De Los Reyes & Kazdin, 2005). In the current study we, for example, asked the students to rate how many positive and negative behavioral responses they had experienced during the last month (i.e., the 30 days prior to assessment). We speculate, that young students may not be particularly sensitive to changes in the *amount* of the behavioral feedback they receive from teachers, and that they

primarily perceive that they are receiving *some* response from teachers, and how this attention make them *feel*. Moreover, younger students may have difficulties in remembering how the situation was several days or weeks ago. Asking them to rate the teachers' disciplinary practices and their relationships in class retrospectively, may represent a measurement problem in the present study. Moreover, the introduction of clearer expectations of how to behave may have felt restrictive to some students, and more frequent positive feedback to all students may have felt unfair, particularly among well-adapted students. Yet, we cannot exclude, that the more positive staff ratings compared to student ratings may have reflected exaggerated expectations of positive change among the staff (i.e., positive response bias). It may also be that the changes in the teachers' practices perhaps were too small to be noticed by the students. Whether or not this was the case could not be clarified by the study, since no observations of individual teacher fidelity were made.

Previous studies have shown that older student ratings can provide reliable and valid information, while there are concerns about the validity and reliability of the ratings of younger students (e.g., De Jong & Westerhof, 2001; Ferguson, 2012; Kunter & Baumert, 2006). These concerns relate to the students ability to discriminate among different facets of teaching and social interactions. Younger students are also unaccustomed to rate teacher behavior, and confounding factors may be teacher popularity and emotional bonds to the teacher (e.g., degree to which the teacher is perceived as fair, supportive, and respectful) (e.g., van der Scheer et al., 2019; Wagner et al., 2016). Student ratings can also be strongly influenced by their workload or achievement pressure, and influenced by recent situations and interactions rather than by the ʻmean' or ʻtypical' classroom situation or teacher behavior (e.g., Griffin, 2004; March & Roche, 2000; van der Scheer et al., 2019; Wagner et al., 2016; Weinstein, 1985). Teachers may also vary considerably in their ability to influence students' school behaviors, to establish positive relationships with students, and in their instructional

and classroom management skills (Backers & Hansen, 2015; Gershenson, 2016; Jackson, 2018; Jennings & DiPrete, 2010; Koedel, 2008; Kraft & Grace, 2016; Ladd & Sorensen, 2015; Ruzek, Domina, Conley, Duncan, & Karabenick, 2015).

Incidentally, discrepancies in teacher and student ratings may be due to volatility of classroom instruction, different reference periods (e.g., last week vs. last month), and differences in items used to measure the students' and teachers' perceptions of the classroom climate and disciplinary practices (e.g., items focusing on the student's perceptions of his relationship to the class teacher vs. the class teacher's perception of the overall student-teacher relationships in class) (Clausen 2002; Curby et al., 2011; Weinstein, 1985). Additionally, low between-rater agreement in longitudinal studies may be a consequence of teacher turnover. For example, a class's main teacher (i.e., the teacher to be rated according to the instruction of the student questionnaire) is the same person in Grades 1-4 while another person may hold this position in Grades 5-6, and may be by a third person in 7th grade.

To sum up, although there is consensus that data should be collected from multiple informants, there is little agreement concerning how to use, weight, and balance different reports from for example, students and teachers. What is clear from prior research, however, is a) that high agreement between teacher and student ratings should not be expected in intervention studies like the present one, b) that significantly lower intervention effect estimates are to be expected in student ratings than in teacher ratings, c) that there may be several reasons, other than measurement error, for why the intervention outcomes seemed less positive in the student ratings than in the staff ratings in the current study, and d) that the observed informant discrepancies do not mean that the effect estimates based on teacher ratings are ˊwrongʼ and effect estimates based on student ratings are ˊrightʼ or vice versa. This review of the research indicates that teacher and student assessments of the learning environment or classroom behavior probably reflect genuinely different perspectives that

should be analyzed separately.

6.6 Comparing Implementation Strategies

Two different implementation strategies were tested in this study. One was the standard or full scale N-PALS model, in which the whole implementation strategy was tested over three years with several strategies for school-wide implementation (each intervention school had a local coach who was trained and certified by NUBU, the coach then trained the school's implementation team, the team teached the rest of the staff, and implementation fidelity was reported annually by all staff). In the short version, the Norwegian PPBS model, the developers trained the whole school staff directly, and no follow up activities like additional training or supervision or coaching were offered. Interestingly, the PPBS outcomes related to school problem behavior (but not on other outcomes) were almost equally positive as the N-PALS outcomes, but only measured in the short run (four months after training). While the full model needed one to three years to become effective, the short version gave immediate positive teacher reported outcomes.

In the current study, the implementation quality was measured with the EBS-SAS self-assessment survey (Sugai & Horner, 2009). This self-report was filled out by all staff once a year, and asked about the current status of different features of the school model, related to the school-wide system, the non-classroom setting system, the classroom system, and individual student system. The current status could be 'in place', 'partial in place' or 'not in place'. The assessment system was primarily addressing whether the core ingredients of the model, according to the individual staff member, had been implemented or not. This could be characterized as a perceived measure of the model implementation, and gave no information about how the individual staff member adhered to the model in practice. It was therefore impossible, based on this assessment, to evaluate the extent to which the individual staff

member practiced the model as intended. In other words, the EBA-SAS assessment was not adequately sensitive to capture variations in the teachers' practices.

Seventy-five percent of the N-PALS intervention schools had implemented the model with required fidelity (i.e., minimum 80% on the overall EBS-SAS scale, Sugai, Horner, & Todd, 2009) as measured after three years. The implementation quality varied, however, across school areas and intervention levels, and also with school size and characteristics of the staff and student body. Strategies and interventions related to the school-wide, common arenas, and the classroom context (i.e., universal level) were well implemented in most schools (86%-96%) whereas strategies and interventions for at-risk students (i.e., selected and indicated level) were only sufficiently implemented in less than a third of the schools. The fidelity scores were significantly higher in smaller schools (n <200 students), b) in schools with low portion of untrained staff, c) low portion of students with immigrant background, d) and less challenging students (e.g., portion of special needs students, portion reported to the police, expelled, and segregated from ordinary class). These results point to that this school-wide framework is challenging and takes time and effort to install or implement with fidelity across all staff members, all classes, and the school context as a whole.

In line with prior research, implementation quality moderated all the primary outcome variables (i.e., variables related to school problem behavior) in the schools implementing the N-PALS model (see Table 6.1). In addition, implementation quality moderated staff-rated classroom climate and student-student relationships in class as rated by the students. In other words, better outcomes were achieved in the high-implementing schools on these variables. Implementation quality did not, however, moderate the outcome variables related to the school staffs' disciplinary practices and self-perceived efficacy. This lack of moderating effects likely indicates that all intervention schools achieved equally good results regardless of their assessment scores. In the evaluation of the PPBS intervention, implementation quality

only moderated the effects on classroom climate as rated by the staff and on collective efficacy. Fidelity in these intervention schools was measured only four months after completion of the program training. We speculate that the lack of moderation effects in this case may be partly due to a too short time-span to achieve enough between-school variation in the implementation scores (a ceiling effect, i.e., high proportion of schools with high fidelity scores).

Prior evaluation studies of the SWPBS model have found that high implementation quality is systematically associated with better outcomes (e.g., Bradshaw et al., 2010; Dix, Slee, Lawson, & Keeves, 2012; Jolivette et al., 2014; McIntosh et al., 2011; Pas & Bradshaw, 2012; Ross, Romer, & Horner, 2012; Sørlie & Ogden, 2007). And generally, incomplete or inaccurate implementation of interventions results in less outcome effectiveness than complete and accurate implementation (e.g., Berkowitz & Bier, 2007; Domitrovich et al., 2008; Durlak & Weissberg, 2007; Durlak et al., 2011). Although high fidelity increases the odds of positive intervention effects, it is not deterministic of program outcomes (Killerby & Dunsmuir, 2018). But results from a recent review of fidelity in 42 K-12 classroom-level effectiveness studies (Hill & Ericson, 2019) showed that interventions implemented with moderate or high fidelity had more than double the chance of achieving positive results than null results. However, Hill & Ericson (2019) also found that six of 42 effectiveness studies reported majority-positive impacts but low fidelity scores, and eight studies reported strong fidelity yet majority-null effects. These findings indicate that implementation fidelity is a partial but not complete explanation for positive intervention effects as well as for null effects.

Several reasons for low implementation fidelity have been discussed, from inadequate principal or colleague support for the intervention, insufficient knowledge among teachers to fully understand and implement the intervention, competing job tasks to that when the new practices and strategies have become routine, inaccuracy likely occur in the school staff's

actions (e.g., Borman, Gamoran, & Bowdon, 2008; Givvin & Santagata, 2011; Santagata, Kersting, Givvin, & Stigler, 2011). Characteristics of the program design (e.g., several versus few components, supervision or not), program materials, and of the context in which the intervention is implemented may also affect fidelity (Jacob, Doolitle, Kemple & Somers, 2019; Kim, 2019). Moreover, interventions requiring more difficult tasks or new practices that teachers previously have been unfamiliar with — for instance, using data to inform instruction and systematic rewards of expected student behavior — may simply be more challenging for teachers to enact and less likely to be implemented with fidelity.

The schools participating in the present effectiveness study were among the very first schools in Norway implementing a school-wide and structured prevention framework. In addition to focusing on core model components and strategies of the N-PALS framework, the training of the implementation teams and the local N-PALS consultants focus on the implementation structure. One may, however, speculate if the training sufficiently emphasizes *why* implementing (or practicing) interventions with high fidelity is important, *why* the schools should regularly monitor the implementation quality, *how* to use implementation data to adjust the implementation process, and *how* they can sustain high implementation quality over time. Moreover, it is uncertain to which degree the N-PALS teams were capable of training their colleagues in the core model and practices, and to what degree the teams in fact did complete the expected school-internal model training. Due to a lack of research, no valid information exists on these areas at present.

6.7 Magnitude of Intervention Effects

A major problem with statistical significance tests is that almost any difference or correlation will be significant if the samples are large enough (e.g., Ferguson, 2009: McCartney & Rosenthal, 2000). There are several ways to help translate results from statistical significance

testing into results that can be practically applied to real life situations, of which effect size statistics, as used in the present study, is the most preferred (Pogow, 2019). A large ES indicates a stronger effect. ES used in effectiveness research is an index that quantifies the amount of difference in pre-post change between groups, and thereby the degree to which the study results should be considered practically negligible or important, regardless of the size of the study sample (Pogrow, 2019). ESs can also be used to quantitatively compare the results of two or more studies (e.g., to decide whether one intervention is more effective than another), the results across variables in a given study (e.g., to decide whether the intervention benefits were larger on behavior outcome variables than on academic outcome variables), and to guide decisions about how many subjects are needed in a future study (power analyses) (Cohen, 1977).

When assessing group differences in the current study, we used Cohen's *d* which is the most commonly used ES measure. Cohen's *d* relies on the pooled standard deviation to standardize the measure of ES. The rule of thumb or benchmark suggested by Cohen (1977) is that a *d* of .20 usually should be considered a small intervention effect, a *d* of .50 is moderate, while a *d* of .80 indicates a large effect. Hence, Eyberg, Nelson, and Boggs (2008) and others (e.g., Hill, Bloom, Black, & Lipsey, 2008) suggested that an ES of .20 should be set as the standard for satisfactory practical or clinical significance. Others have discussed Cohen's *d* in relation to educational interventions. Lipsey et al. (2012) underscore that Cohen's broad categories of small, medium, and large are not tailored to the effects of intervention studies in education, and "Using those categories to characterize effect sizes from education studies, therefore, can be quite misleading" (p. 4). To exemplify, they note that many school intervention studies investigate effects on academic performance as measured with standardized reading or math achievement tests. However, ESs found across a wide range of interventions are seldom as large as .30. By using more appropriate norms, i.e., norms based

on empirical distributions of ESs from comparable studies, “an effect size of .25 on such outcome measures is large and an effect size of .50, which would be only ‘medium’ on Cohen’s distribution, would be more like ‘huge’ (Lipsey et al., 2012, p. 4). Kraft (2019), who reviewed 481 ESs from 242 RCTs, suggested that in education, ESs < 0.05 should be considered small, 0.05 < 0.20 medium, and 0.20 or greater as large.

According to Thompson (2004), ES for all results should be reported, including those that are statistically non-significant. Thompson (2004) holds that not reporting ES for non-significant results is the same as treating these effect sizes as zero. Some also have argued that to help researchers and practitioners to better determine the importance of a result, the confidence intervals (CIs) of the ESs also should be given (e.g., Ialongo, 2016; Thompson, 2004). CI is a way of indicating the margin of error of a finding. Confidence intervals thus tell us about the precision of the effect estimates. The narrower the confidence interval the more precise is the estimate and we have a high degree of assurance that the results are true or nearly true. In the current study, the ES and 95% CI for all outcomes are reported.

Several have warned that ES estimates from single studies and Cohen’s *d* should not be applied rigidly when deciding upon the practical significance of results and when determining the effectiveness of interventions for improving student or client outcomes, and/or teacher, parent, and therapist practices (e.g., Ferguson, 2009; Hill et al., 2008; Ialongo, 2016; McCartney & Rosenthal, 2000; Pogow, 2019; Prentice & Miller, 1992; Sink & Stroh, 2006, Thompson, 2002, 2004). Objections vary from that a) effect sizes observed in social sciences are oftentimes very small and that this has led to difficulties on their interpretation (e.g., Rosnow & Rosenthal, 2003), to b) that ESs just as children are best understood in context (e.g., Hill et al., 2008; McCartney & Rosenthal, 2000; Thompson, 2002, 2004; Sink & Stroh, 2006), c) that the minimum effect size cut-off of representing a practically significant effect perhaps is set too low (Pogow, 2019, referring to intervention effects on academic

achievement only) or d) too high (Lipsey, 2000), and to e) that sometimes small ESs are impressive or should not be erroneously dismissed as trivial (e.g., McCartney & Rosenthal, 2000; Prentice & Miller, 1992). Examples of the last objection can be a) when an intervention results in a lowered death rate of say 1%, b) when an universal intervention results in a small overall effect (impact on many persons) - in particular, if other interventions exist in parallel to the one being evaluated, c) if the intervention implies relatively low costs and minor interruptions of daily practice, or d) when an intervention has no discernable impact in regard to making the intervention group indistinguishable from the control group, but still may do a lot to help people cope with their problems or symptoms and/or to improve their perceived efficacy or quality of life.

Although an intervention effect in standard units is not influenced by sample size and provides more informative data than statistical significance in evaluating the practical importance of findings, there still is a risk for overinterpretation of small effect sizes due to that the ES can be biased by other factors, such as measurement, design, and methodological choices. Better measurement, design, and method all tend to produce larger effects. Interpretation of effect sizes thus should be context specific. Accordingly, and in line with McCartney and Rosenthal (2000) and others (e.g., Thompson, 2004) suggestion, the effect estimates and practical significance of the current study should be evaluated in relation to a scientific and empirical literature context.

6.7.3 Effect sizes in the current study. The effect size estimates reported in Paper 2-4 are summarized in Table 6.2. The ESs are the differential change between pretest and posttest. In other words, by showing effects as differential change, we not only reveal the end result but the patterns of improvement or decline that characterize the interventions and control group (Lipsey et al., 2012).

Table 6.2. Summary of Effect Sizes with 95% CI for N-PALS and PPBS

Outcome variables	N-PALS (after 3 years)		PPBS (after 4 months)	
	d	95% CI	*d*	95% CI
Problem Behavior in the School Environment, teacher rated	0.25	0.236 - 0.254	0.38	0.375 - 0.385
- moderate	0.24	0.236 - 0.244	0.41	0.405 - 0.415
- severe	0.17	0.166 - 0.174	0.18	0.175 - 0.185
Problem Behavior in the Classroom Context, teacher rated	0.13	0.126 - 0.133	0.13	0.125 - 0.135
- moderate	0.12	0.116 - 0.124	0.11	0.105 - 0.115
- severe	0.11	0.106 - 0.114	0.15	0.145 - 0.155
Classroom Climate, teacher rated	0.17	0.166 - 0.174	0.08	0.066 - 0.076
Classroom Climate, student rated	0.01	0.009 - 0.011	0.04	0.039 - 0.040
- student - student relationships	0.10	0.099 - 0.101	-	-
- student teacher relationships	0.00	0.001 - 0.001	-	-
Teacher Collective Efficacy	0.34	0.336 - 0.344	0.12	0.124 - 0.116
Teacher Self-Efficacy	0.14	0.126 - 0.144	0.10	0.104 - 0.096
Positive Behavior Support, teacher rated	0.91	0.906 - 0.914	0.41	0.019 - 0.021
Positive Behavior Support, student rated	0.01	0.009 - 0.011	0.02	0.094 - 0.086
Behavioral Corrections, teacher rated	0.01	0.006 - 0.014	0.09	0.019 - 0.021
Behavioral Corrections, student rated	0.05	0.049 - 0.051	0.02	0.124 - 0.116

From Cohen's benchmark indicators, the significant effect sizes for the two intervention frameworks ranged from relatively small-modest to large. According to Lipsey et al. (2012) and Kraft (2019) they should be considered moderate to large. The confidence intervals were narrow, indicating high precision of the estimates.

The school level effect size estimates related to N-PALS and PPBS's primary goal (reduce prevalence of problem behaviors in school), compare well with those reported in a) prior rigorous effectiveness studies of SWPBS (e.g., Bradshaw et al., 2012) b) meta-analyses and systematic reviews of SWPBS outcomes (Gage, Whitford, & Katsiyannis, 2018; Horner, Sugai, & Anderson (2010); Lee & Gage, 2020; Noltemeyer et al., 2019: Solomon, Klein, Hinze, Cressey, & Peller, 2012; Stewart, Benner, Martella, & Marchand-Martella, 2007), and c) meta-analyses of universal school-based interventions (DuBois, Holloway, Valentine, & Cooper 2002; Durlak et al., 2011; Hahn et al., 2007; Weare & Nind, 2011; Wilson & Lipsey (2007); Wilson, Lipsey, & Derzon, 2003). The ES estimates being comparable to other

intervention studies and meta-analyses support the reliability and validity of our findings.

The large scale of the study, the strengthened non-randomized design, together with that most intervention schools implemented with acceptable fidelity, the narrow CI intervals, and that the ES estimates matched well with those reported in other SWPBS studies, research reviews, and meta-analyses make us suggest the study outcomes being sufficiently accurate, robust, meaningful, and large enough to be of practical significance. However, although well established and reliable measures were preferred in the current study, we cannot exclude that some measurement error and influence from unobserved third-variables may have biased the effect estimates downwards.

6.8 Study Strengths and Limitations

The current study has several considerable strengths, still the findings should be considered in light of some potentially notable methodological limitations related to the non-randomized design and some of the measures used.

A strength is the testing of intervention effects among normally developing children and within ordinary school settings. Other advantages are that the interventions could be implemented without violating the national curriculum nor the norm of inclusion, and within the schools' ordinary budget with minimal extra costs. Compared to most prior school-based effectiveness studies, a rather large sample of schools, high response rates across assessment points, and the use of multiple informant groups, including the entire school staff, represent other study strengths. Multilevel analyses have several benefits when school-based intervention effects are to be studied, including accounting for nesting in data, correcting for pre-test scores, and for influence on the effect estimates stemming from other factors than the intervention. Potentially confounding factors accounted for in the outcome analyses (i.e., covariates) of N-PALS and PPBS were school size, proportion of staff members without

formal training, portion of staff working in the after school services, proportion of special education students, and proportion of students with immigrant background. Accordingly, the effect estimates were robust even when controlling for initial school variance and several other possibly influencing school level factors beside the intervention itself, including significant characteristics of the staff and student body. Although, some potentially confounding factors were controlled for, there may be other unmeasured factors that can have influenced the results (e.g., change of principal during the study period, the existence of school differences more than one school-year prior to the initiation of the study).

The choice of outcome variables included in this dissertation was closely related to the main purpose and goals of the standard N-PALS model and the PPBS intervention, and as far as possible, the key constructs were measured at the appropriate unit of analysis. In evaluating the effects of the school-wide interventions, we started with the estimation of school-level effects rather than individual- or class-level effects. Analyses of secondary outcomes, such as changes in the students' academic performance, social skills, and motivation, and differential outcomes for sub-groups of students and potential intervention ´spill-over' effects to the home context are not a part of the thesis, but are planned published in later papers.

A limitation relates to the (partly) non-randomized study design. Too few schools accepted to participate in the initially planned true experimental design (RCT), and instead a strengthened non-equivalent cluster-based comparison group design (NEC) was applied in order to reduce as many plausible threats to the internal and external validity as possible. The schools were randomly selected to participate in the study, but they were not randomly assigned to the intervention or control groups. A combined stratifying and matching procedure was, however, applied to make the intervention and comparison groups initially comparable. Very few significant group differences were found at baseline. Although several measures were taken in order to match the intervention and comparison group on key

variables and to secure that the participating schools were representative of Norwegian schools, there may still be, as noted by Shadish, Cook, and Campbell (2002), that even the best designed non-randomized experiments cannot rule out the chances of differences in unmeasured variables related to the outcomes.

Not all schools invited to participate in the study as intervention or control schools accepted the invitation, but any selection bias may have been similar in both groups in that no differences were found between the declining and accepting group. As accounted for in Paper 2-4, the attrition in the current study was relatively high. We do not, however, see this as a serious threat to the validity because this could be expected from the open cohort design and normal fluctuations in the staff and student population across time. No systematic differences in attrition between the intervention and control group were observed.

Although, we cannot completely rule out that some selection bias may have occurred in the process of recruiting schools to the study or that all other threats to the validity were fully eliminated, we nevertheless assert that there is a reasonable basis for claiming that the strengthened NEC-design together with the statistical control for initial school variance and potential confounders in the outcome analyses, contributed to high statistical power and valid effect estimates, in particular those based on staff ratings.

The current study was not designed to identify the relative importance of each component that was a part of the multilevel, multicomponent model or framework. It is a structured approach addressing three levels or tiers in school, the universal, the targeted and the indicated level. Being both a prevention and an intervention model, it may be considered a particularly complex and difficult model to implement, and also to evaluate. The analyses and outcomes reflect that it was the whole model that was evaluated and produced positive outcomes.

The school staffs in the N-PALS and PPBS groups were responsible for implementing

the interventions, and also reported on most outcomes and implementation quality. Although, we cannot reject that evaluating one's own efforts may lead to biased estimates, we do not believe this represents a major problem in the current study. First, any positive response bias likely has affected both the intervention and the comparison group equally (remind that despite implementation of other programs was an exclusion criteria, most control schools in fact implemented other programs during the study period). Second, participating in the study should be of equal interest for both groups because the school staff respondents are probably the stakeholders most concerned about the challenges of student problem behavior in school. Third, the current design with repeated measures over several years reduced the likelihood of any Hawthorne or Novelty effect causing positively skewed effect estimates.

Another limitation may relate to that EBS-SAS (see Paper 2-4) was used as the sole implementation measure. Although several studies have used this scale as an implementation measure, it was originally developed for needs assessment, not specifically as a fidelity measure. The implementation measurement asked staff if several of the model components were installed and practiced at their school, but it did not measure staff fidelity in daily practice; for instance, how frequently they used praise and encouragements to promote positive student behavior. Indeed, the staff members were asked about their behavior management strategies, but these were summary self-report statements and not related to actual lessons. Through the moderation analyses we were able to indicate the impact of some school contextual variables (e.g., implementation quality, school size, proportion of untrained staff). However, the moderators were not tested in separate studies in order to assess their causal inference.

Limitations of to our analyses of the relationship between perceived teacher collective efficacy and problem behavior in school are that the conclusions relate to cross-level main effects, a relatively short time span between measurements, and the observed reciprocity

between the two phenomena give reason to emphasize that the deterrent influence of collective efficacy on school problem behavior is tentative only, and that the potential effects of collective efficacy on individual student behavior were not explored.

The lack of coherence between staff- and student-based effect estimates may represent another limitation. The students did not report any improvements in the intervention schools as compared to student reports in the control schools. As discussed in section 6.5, this raises several questions. It may be that the students did not approve the changes following the intervention models, or did not perceive or notice anything different in the teachers' attitudes or behavior. Maybe most students did not feel they needed the extra reminders, teaching of rules, or positive behavior supports? It may also be that the teachers did not improve their interactions with the students, and that the positive changes reported were colored by their expectations. Another possibility is that the measurement failed, for instance, by being too general and requiring the students to be more observant than could be expected from students at this age. If students had been asked other questions or maybe interviewed, more detailed feedback could have been obtained. Moreover, if the study had included direct observation, it would also have been easier to decide what the reasons for the lack of positive student-rated outcomes were. However, additional interviews or observations were not possible due to a lack of resources.

A variable examined in many implementation studies is the one called 'readiness for change', which signals to what extent the practitioners have a high problem awareness and are motivated to participate in projects aiming at changing their practice. Although not reported in the papers included in the dissertation, readiness for implementing N-PALS (p = .006) and the school's perceived need for change (p = .000) were rated by the principals, indicating significantly greater needs and readiness in the intervention schools as compared to the comparison schools. Still, the experiences and outcomes of this project make us speculate if

some of the participating schools and teachers were adequately prepared for an intervention of this magnitude and structure. Part of the unpreparedness may relate to the increased demands and expectations to strengthen the infrastructure of the school with several new routines, activities, written guidelines, and the expectation that the implementation team as well as the staff should be accountable for their activities.

Lastly, the study outcomes also showed considerable positive changes over time in the comparison schools, which likely reflect that these schools had implemented other effective (although less comprehensive) school programs, and/or that the general practice in Norwegian primary schools holds a reasonable high quality.

6.9 Conclusions

The N-PALS model (normal and abbreviated version) was developed and implemented in order to prevent and reduce externalizing student problem behavior. The behavioral problems are not considered an expression of negative intentions or the desire to disturb or destroy, but as a problem of paying attention, managing emotions, completing tasks, and verbally communicating their needs. What they need are skills which help them to be respectful, control their impulses, wait their turn, stay focused on their work, and remember rules and expectations. If only one, or a few students lack these skills, the whole classroom or school may turn into chaos. Therefore, N-PALS is a collective enterprise in which all staff encourages and models to all students the core expectations, rules, and skills necessary to develop a safe and productive learning environment. The background and aims of the N-PALS model and this study are to contribute to the prevention and reduction of the amount of student behavior problems in Norwegian primary schools. The school-wide model targets the school behavior of the students, with emphasis on externalizing problem behaviors, which are among the most ‘wicked problems’ in our schools. According to teachers, students with

behavior problems are among those most difficult to include in schools, and the group most often referred to alternative learning arenas. The challenges of students and school classes out of control, tend to increase as the students get older. Therefore, it is of great importance to find out if these problems can be addressed and prevented by analyzing the problems at an early stage.

Without doubt, some students are in need of a more structured learning environment in which they are taught clear rules and expectations, receive more positive behavior support, and mild behavioral corrections. Students with self-regulation deficits and attention problems may particularly profit form structured education and a structured school environment with increased amount of supervision. However, students who don't have such problems or short-comings, could probably benefit from a less structured environment in which they may practice self-regulation and planning skills. The N-PALS model therefore intends to strike a good balance between the students' different needs by increasing the social and educational structure of the school, while at the same time give students who adapt well and achieve to their potential the positive feedback and encouragement they are entitled to.

As measured after three years of implementation, the N-PALS model seemed effective in reducing the level of both moderate and severe externalizing student problem behavior (particularly on common school areas) and in promoting qualitatively better social and academic learning conditions in class. The results also indicate increased ability in the intervention schools to reach out to all students and reduce the number of students singled out for individual intervention plans due to conduct problems.

Moreover, the SWPBS model as implemented in Norway appeared effective in promoting school staffs' perceived individual and collective efficacy. Besides, this three-level prevention framework appeared effective in changing staff behavior in positive ways. Opposite to the control group, the members in the N-PALS schools substantially increased

their use of positive behavior supports over time. A co-occurring decrease (even non-significant) in behavioral corrections was observed in the N-PALS schools, and they reached the suggested ideal ratio of 4-5:1 between support and correction.

Furthermore, the outcome analyses of the abbreviated version of N-PALS, the PPBS intervention, indicated promising, immediate effects on several but not on all of the same outcome variables as the full three-level model. These results still make us argue that for some schools the PPBS intervention may be appropriate as a basic approach or a ʻspringboardʼ to facilitate the development of school-wide, more common attitudes and strategies related to the prevention of behavioral misconduct.

Looking back, it is natural to ask questions about what we have learned from this study besides that a school-wide and structured approach to prevent school problem behavior and to promote a positive learning environment for all students is both effective and feasible. In line with other comprehensive and multi-faceted studies, the study has created as many new research questions as it has answered. Evaluated by any standard, it is evident that the model has contributed positively to the development of inclusive schools. The outcomes are not entirely in favor of the interventions, and most effect sizes are relatively modest. We take this as a positive sign because most comparable studies show effectiveness in the same range as this one, and we believe that schools change in an incremental way.

One of the first learning experiences was that it was extremely difficult to recruit schools for a randomized controlled trial without wait-list control. The possibility of being randomized to the control condition for the entire study period did not appeal to most schools. We also learned that implementation matters – because implementation scores moderated several outcomes in the intervention schools. Schools with high implementation quality scores benefitted most from the intervention. Other studies have shown that the implementation of the SWPBS model is difficult, indicating that 50-70% of the schools who have adopted this

prevention framework have implemented it with satisfactory fidelity. Furthermore, we learned that school-size matters - because small and medium sized schools had more positive outcomes compared to large schools. Incidentally, we have learned that professional training matters - because schools with a high mean training dosage and a high portion of staff without professional training benefitted most from the intervention.

The current study adds to the existing base of knowledge in several ways, including by being the first to examine the cumulative effects of the three-level model and by being the first effectiveness study of the SWPBS model implemented in Europe and in a non-English speaking context. Moreover, the Norwegian study is the first investigating effects of SWPBS on perceived teacher collective efficacy and on school staffs' disciplinary practices, and also one of very few longitudinal effectiveness studies to date that have been conducted on the SWPBS model. Besides, our study is the first to investigate the relationship between collective efficacy and problem behavior in school, and on an empirical basis suggesting that teacher collective efficacy is a significant and reliable indicator of a school's culture.

The Norwegian outcomes indicate that the SWPBS model was successfully transferred across the Atlantic without major adaptations, at least to Norway as a representative of the Scandinavian and west-European countries. Ideally, the study results should be replicated in a Scandinavian randomized controlled trial. However, a recent systematic review of the existing research published in the journal Ungsinn (Asheim, Patras, Eng, & Natvig, 2018) concluded that the N-PALS model should be classified at the highest evidence level (5), that is, as a ˙measure with strong documentation of effect' – and with that, an intervention model ready for broad dissemination.

6. 10 Implications for Future Research

In the analyses included in the dissertation, the N-PALS and PPBS interventions were

compared to the control group but not to each other. Accordingly, the three-group quasi-experimental design was not exploited to the fullest. In future analyses based on data from the current effectiveness study, this should be done. Moreover, the follow-up assessment in the N-PALS (end of year four) and control schools was not used. Likewise, in future analyses we have the opportunity to evaluate potential longitudinal effects of PPBS, and as noted earlier, examine if one or both interventions might have impacts on more distal (secondary) outcome variables. Besides, future efforts should be made in validating several of the measuring instruments used, particularly those developed for the study.

For the time being, research lacks on several other important areas related to the N-PALS model; for example, whether some model components are more influential than others, which are the operating processes that induce changes in student behavior and skills (e.g., through changes in the staffs behavior and/or attitudes?), and how can high implementation quality best be sustained in schools over time (e.g., through changes in the structure of the training, the school feedback system or local supervision?). Likewise, there is a lack of research on intervention outcomes for different sub-groups of students, such as for students at respectively moderate and high risk of conduct disorder, special education needs students, students with immigrant background, boys and girls, and for students in need of support from the child welfare system. Moreover, and despite the documented positive school level effects in elementary schools and the highest classification of evidence, we still do not know whether the intervention effects are transferrable to kindergartens, middle schools, and high schools due a lack of political support and economic resources to extend the implementation of N-PALS.

6.11 A School Model of The Future?

The dissertation closes with some final thoughts about the N-PALS model and the way ahead.

Currently, the SWPBS/N-PALS model belongs to the group of promising approaches that may be used by schools to induce powerful and sustained school-wide changes to the benefit of the staff and students. Sugai and Horner (2002) wrote:

> Schools are learning that the «tricks» of behavior and classroom management are insufficient to achieve meaningful and sustained improvements in student behavior, especially when the problem behaviors are chronic and intense. More importantly, the behavior management capacity of the school is not enhanced when the focus is on reacting to one situation at a time. (p. 30).

Still, there is room for improvements both of the model itself and the implementation strategies. For example, the N-PALS model could be further developed into the ideal Multi-Tiered System of Support model (MTSS, see section 2.4.2). This extended version, which focus both on preventing behavior problems, social skills deficits, and academic problems, has been described as I-PALS in Norway (Arnesen & Meek-Hansen, 2011). As a first step in this process, reliable and valid measurement scales for assessing early reading problems and social skills among elementary school children have been developed (Arnesen et al., 2016; Arnesen, Smollkowski, Ogden, & Melby-Lervåg, 2017). A natural second step would be to pilot and add evidence-based interventions in order to promote the students˙ academic and social skills. A third step would be to ensure relevant training and supervision in this extended intervention model together with a refined quality assurance system. A fourth step would be to conduct a new high-quality effectiveness study.

Moreover, in order to increase and better support the schools' implementation knowledge and implementation fidelity over time, the model adapters/purveyor should discuss the need for more focus on these themes during the training, and for more frequent follow-up from the local consultant during the supervision of the N-PALS teams. It also should be discussed whether an extension or refinement of the existing training structure of the N-PALS model in the direction of the one used in the PPBS intervention may be appropriate (e.g., in addition to team-training, offer some additional whole-staff training).

Despite that some refinements to the model and implementation structure should be considered, the N-PALS framework represents a promising solution to several schools challenges emphasized in recent political documents. For example, the Report to the Storting (i.e., the Norwegian Parliament) no. 21 (2016-2017) and no. 6 (2019-2020) from the Ministry of Education and Research, emphasize the need for an increased focus on an inclusive school community, early interventions, greater local room for quality development, and a durable raise in the competence among teachers and the special education needs staff in kindergartens, schools, and the Educational-Psychological Service. Besides, a Scandinavian expert group (Nordahl et al., 2018) asserts that the supports offered to Norwegian students in need of special education are initiated too late, and with too low quality and effectiveness. The expert group suggests (p. 8-9, translated by author) that; a) "All children and youths in need of adapted teaching in kindergartens and schools should be helped and supported wherever they are", b) "This help and support must be effectuated early, be individually adapted, and take place within an inclusive school community", and c) "All children and youths in need of specially adapted teaching shall meet teachers with relevant and formal educational competence". The existing Norwegian and international evidence on the N-PALS/SWPBSs impacts indicate that these needs by far can be met by implementing this school-wide, preventive, and multi-level model.

Moreover, a recent Official Report (NOU 3:2019) from the Stoltenberg committee and a report from OECD (2018) focusing on the rather large achievement gender-gap disfavoring boys in Norwegian schools, both explicitly suggest, with reference to existing research and experiences from among others the USA, Finland, England, and the Netherlands, that the N-PALS preventive system framework should be implemented on a large scale in our schools.

Together with other measures, such as providing teachers with concrete tools and guidelines on how to address students' behavioral issues ping out, and encouraging schools to

introduce tutoring and mentoring in the classroom, providing schools with standardized tools to identify students at risk of dropout, and programs targeting low achieving boys, OECD (2018, p. 106-107) suggests that Norway should consider ˙Testing and evaluating the impact of a multi-tiered support system on improving students learning outcomes' and argue that ˙Such model can be a good alternative to Norway's current SEN (author: special education needs) model and can provide support to a bigger number of low achieving boys at risk of dropping out'. The Stoltenberg committee argues that according to the Education Act (§ 9a), the school must actively and systematically promote a good psychosocial environment, in which the individual student can experience safety and social belonging (NOU 3:2019, p. 151), which are at the core of the N-PALS model (ibid, p. 153). The committee adds that the school, eventually also the kindergarten, are suitable as bases for preventive and health-promoting measures such as the N-PALS model because they reach virtually all children and families, and preventive efforts can be carried out without segregating at-risk students. Additionally, several preventive efforts, including the N-PALS model, focus on skill deficiencies where boys are overrepresented, such as violence, crime, substance abuse, health problems, and lack of social and academic skills (ibid, p. 151).

References

A

Aarts, E., Verhage, M., Veenvliet, J. V., Dolan, C. V., van der Sluis, S. (2014). A solution to dependency: Using multilevel analysis to accomodate nested data. *Nature Neuroscience, 17,* 491-496.

Abidin, R. R., Greene, R. W., & Konold, T. R. (2004). *ITS, Index of Teaching Stress: Professional Manual*. Psychological Assessment Resources.

Achenbach, T. A. (2011). Commentary: Definitely more than measurement error: But how should we understand and deal with informant discrepancies. *Journal of Clinical Child & Adolescent Psychology, 40,* 80-88.

Adi, Y., Killoran, A., Janmohamed, K., & Stewart-Brown, S. (2007) *Systematic review of the effectiveness of interventions to promote mental wellbeing in primary schools: universal approaches which do not focus on violence or bullying*. London: National Institute for Clinical Excellence.

Alquraini, T., & Gut, D. (2012). Critical components of successful inclusion of students with severe disabilities: Literature review. *International Journal of Special Education, 27,* 42-59.

APA (2009). *Publication manual for the American Psychological Association (6th ed.).* Washington, DC: American Psychological Association.

APA (American Psychiatric Association) (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®), Fifth Edition*. Arlington, VA: American Psychiatric Association.

Arnesen, A. (2011). Innsats mot mobbing i skolen. Virker-virker ikke? Kronikk. *www.forebygging.no.*

Arnesen, A., Braeken, J., Baker, S., Meek-Hansen, W., Ogden, T., & Melby-Lervåg, M. (2016). Growth in oral reading fluency in a semitransparent orthography. Current and predictive relations with reading proficiency in Norwegian, grades 2-5. *Reading Research Quarterly*, 0, 1–25.

Arnesen, A., & Meek-Hansen, W. (2010). *PALS Håndbok modul 1. Skoleomfattende forebyggende tiltaksnivå.* Oslo: Norsk senter for studier av problematferd og innovativ praksis.

Arnesen, A., & Meek-Hansen, W. (2011). En integrert skoleomfattende tiltaksmodell for å fremme god leseutvikling og sosiale ferdigheter. *Spesialpedagogikk, 8,* 4-14.

Arnesen, A., Meek-Hansen, W., & Ogden, T. (2011). Konklusjoner på tynt grunnlag. Kronikk. *Dagsavisen* (mars).

Arnesen, A., Meek-Hansen, W., Ogden, T., & Sørlie, M-A. (2014). *Positiv læringsstøtte. Hele skolen med!* Oslo: Universitetsforlaget.

Arnesen, A., & Ogden, T. (2006). Skoleomfattende kartlegging av elevatferd. *Spesialpedagogikk, 2,* 18-29.

Arnesen, A., Ogden, T., & Sørlie, M-A. (2006). *Positiv elevatferd og støttende læringsmiljø i skolen*. Oslo: Universitetsforlaget.

Arnesen, A., Smollkowski, K., Ogden, T., & Melby-Lervåg, M. (2017). Validation of the elementary social behavior assessment. Teacher ratings of students' social skills to Norwegian, grades 1-6. *Emotional and Behavioural Difficulties, 0,* 1-16.

Arnesen, A., Ogden, T., & Sørlie, M-A. (2006). Positiv atferd og støttende læringsmiljø i skolen. Oslo: Universitesforlaget.

Arnesen, A. & Sørlie, M-A. (2012). Forebyggende arbeid i skolen. I E. Befring, I. Frønes, & M-A. Sørlie, *Sårbare unge. Nye perspektiver og tilnærminger* (s. 86-102). Oslo: Gyldendal Akademisk.

Algozzine, B., Daunic,A. P., & Smith, S. W. (2010). *Preventing problem behaviors:Schoolwide programs and classroom practices*. CA, Thousand Oaces: Corwin.

Asheim, M., Patras, J., Eng, H., & Natvig, H. (2018). Kunnskapsoppsummering og klassifisering av tiltaket PALS – Positiv atferd, støttende læringsmiljø og samhandling (2.utg.). *Ungsinn,* 2:1. https://ungsinn.no/post_tiltak_arkiv/palspositiv-atferd-stottende-laeringsmiljo-2-utg/

Astor, R. A., Guerra, N., & Van Acker, R. (2010). How can we improve school safety research? *Educational Researcher, 39*, 69-78.

Atkins, D. C. (2005). Using multilevel models to analyze couple and family treatment data: Basic and advanced issues. *Journal of Family Psychology, 19*, 98-110.

Attar-Schwartz, S. (2009). Peer sexual harassment victimization at school. The roles of student characteristics, cultural affiliation, and school factors. *American Journal of Orthopsychiatry, 79*, 407-420.

Augimeri, L. K, Farrington, D. P., Koegl, C. J., & Day, D.M. (2007). The SNAP™ Under 12 Outreach Project: Effects of a community based program for children with conduct problems. *Journal of Child and Family Studies, 16*, 799–807.

Augimeri, L. K, Walsh, M., & Slater, N. (2011). Rolling out SNAP® an evidence-based intervention: A summary of implementation, evaluation and research. *International Journal of Child, Youth and Family Studies, 2*, 330–352.

B

Backers, B., & Hansen, M. (2015). *Teach for America impact estimates on nontested student outcomes* (Working paper 146). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Bandura, A. (1977). *Social learning theory*. New York: General Learning Press.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 191-215.

Bandura, A. (1997). Social learning theory of aggression. *Journal of Communication, 28*, 12-29.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1, 91-97.*

Baldwin, S. A., Murray, D. M., & Shadish, W. R. (2005). Empirically supported treatments or type I errors? Problems with the analysis of data from group-administered treatments. *Journal of Consulting and Clinical Psychology, 73*, 924-935.

Batsche, G., Elliott, J., Graden, J. L., Grimes, J., Kovaleski, J. F., Prasse, D., et al. (2005). *Response to intervention: Policy considerations and implementation*. Alexandria, VA: National Association of State Directors of Special Education, Inc.

Beelmann, A., & Losel, F. (2006). Child social skills training in developmental crime prevention: effects on anti-social behavior and social competences. *Psciothema, 18,* 603–610.

Berkowitz, M. W., & Bier, M. C. (2007). What works in character education? *Journal of Research in Character Education, 5,* 29–48.

Berry, D., & Connor, E. (2010). Behavioral risk, teacher-child relationships, and social skills development across middle school: A child-by-environment analysis of change. *Journal of Applied Developmental Psychology, 31,* 1-14.

Beunckens, C., Molenberghs, G., & Kenward, M. G. (2005). Direct likelihood analysis versus simple forms of imputation for missing data in randomized clinical trials. *Clinical Trials, 5,* 379-386.

Bickel, R. (2007). *Multilevel analysis for applied research. It's just about regression*. New York: The Guilford Press.

Biglan, A. (1995). Translating what we know about the context of antisocial behavior into a lower prevalence of such behavior. *Journal of Applied Behavior Analysis, 28,* 479-492.

Birnbaum, A. S., Lytle, L. A., Hannan, P. J., Murray, D. M., Perry, C. L., & Forster, J. L. (2003). School functioning and violent behavior among young adolescents: a contextual analysis. *Health Education Research, 3,* 389-403.

Birkett, M., Espelage, D. L., & Koenig, B. W. (2009). LGB and questioning students in schools. The moderating effects of homophobic bullying and school climate on negative outcomes. *Journal of Youth & Adolescence, 38,* 989-1000.

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness, 1,* 289-328.

Bohanon, H., Fenning, P., Carney, K., Minnis, M., Anderson-Harriss, S., Moroz, K., Hicks, K., Kasper, B., Culos, C., Sailor, W., & Piggott, T. (2006). School-wide application of positive behavior support in an urban high school: A case study. *Journal of Positive Behavior Interventions, 8,* 131-145.

Borgen, N. T., Kirkebøen, L. J., Ogden, T., Raaum, O., & Sørlie, M-A. (2019). Impacts of school-wide positive behaviour support: Results from national longitudinal register data. *International Journal of Psychology,*.Doi: 10.1002/ijop.12575

Borman, G. D., Gamoran, A., & Bowdon, J. (2008). A randomized trial of teacher development in elementary science: First-year achievement effects. *Journal of Research on Educational Effectiveness, 1,* 237–264

Bradshaw, C., Koth, C., Bevans, K., Ialongo, N., & Leaf, P. (2008). The impact of school-wide positive behavioral interventions and supports (PBIS) on the organizational health of elementary schools. *School Psychology Quarterly, 23,* 462-473.

Bradshaw, C., Koth, C., Thornton, L., & Leaf, P. (2009).Altering school climate through school-wide positive behavioral interventions and supports. Findings from a group-randomized effectiveness trial. *Prevention Science, 10,* 100-115.

Bradshaw, C. P., Mitchell, M. M., & Leaf, P. J. (2010). Examining the effects of schoolwide positive behavioral interventions and supports on student outcomes: Results from a randomized controlled effectiveness trial in elementary schools. *Journal of Positive Behavior Interventions, 12,* 133–148.

Bradshaw, C. P., Pas, E. T., Goldweber, A., Rosenberg, M. S., & Leaf, P. L. (2012). Integrating school-wide positive behavioral interventions and supports with tier 2 coaching to student support teams. The PBISplus model. *Advances in School Mental Health Promotion, 5,* 177-193.

Bradshaw, C. P., Reinke, W., Brown, L., Bevans, K., & Leaf, P. (2008). Implementation of school-wide positive behavioral interventions and supports (PBIS) in elementary schools. Observations from a randomized trial. *Education and Treatment of Children, 31,* 1-26.

Bradshaw, C. P., Wassdorp, T. E., & Leaf, P. J. (2012). Effects of school-wide positive behavioral interventions and supports on child behavior problems. *Pediatrics, 130,* 1136-1145.

Brandt, R. C., Chitiyo, M., & May, M. E. (2014). Measures used in assessing outcomes of School-Wide Positive Behavior Support, *Journal of Research in Special Education Needs, 14*, 229-238.

Brezina, T., Piquero, A. R., & Mazerolle, P. (2001). Student anger and aggressive behavior in school. A initial test of Agnew's macro-level strain theory. *Journal of Research on Crime and Delinquency, 38*, 362-386.

Brock, L. L., Nishida, T. K., Chiong, C., Grimm, K. J., & Rimm-Kaufman, S. E. (2008). Children's perceptions of the classroom environment and social and academic performance: A longitudinal analysis of the contribution of the *Responsive Classroom* approach. *Journal of School Psychology, 46,* 129-149.

Broidy, L. M., Nagin, D.S.,Tremblay, R. E., Bates, J. E., Brame, B., Dodge, K. A., Fergusson, D., Horwood, J. L., Loeber, R., Laird, R., Lynam, D. R., Moffitt, T. E., Pettit, G. S.,Vitaro, F. (2003). Developmental trajectories of childhood disruptive behaviors and adolescent delinquency: A six-site, cross-national study. *Developmental Psychology, 39,* 222-245.

Bronfenbrenner, U. (1979). *The ecology of human development. Experiments by nature and design*. Cambridge, MA: Harvard University Press.

Bronfenbrenner, U., & Morris, P. A. (2007). The bioecological model of human development. In R. M. Lerner (Ed.), *Handbook of Child Psychology* (Vol.1, pp 793- 828). New York: John Wiley & Sons, Inc.

Brophy-Herb, H. E.., Lee, R. E., Nievar, M. A., & Stollak, G. (2007). Pre-schoolers social competence: Relationships to family characteristics, teacher behaviors and classroom climate. *Journal of Applied Developmental Psychology, 28*, 134-148.

Brown, J. L., Jones, S. M., LaRusso, M. D., & Aber, J. L. (2010). Improving classroom quality. Teacher influences and experimental impacts of the 4rs program. *Journal of Educational Psychology, 102*, 153-167.

Brown, F., Michaels, C. A., Oliva, C. M., & Woolf, S. B. (2008). Personal paradigm shifts among ABA and PBS experts: Comparisons in treatment acceptability. *Journal of Positive Behavior Interventions, 10*, 212–227.

Browne, G., Gafni, A., Roberts, J., Byrne, C., & Majumdar, G. (2004). Effective/efficient mental health programs for school-age children: a synthesis of reviews. *Social Science and Medicine, 58,* 1367–1384.

Budstikka (2014). *Diskusjon om skolepedagogikk: Gir ut «bra-kort»*. 24.01.14.

Bukowski, W. J. (1991). A framework for drug abuse prevention research: Methodological issues. http://www.nida.nih.gov/pdf/monographs/107.pdf

Burke, M. D., Hagan-Burke, S., & Sugai, G. (2003). The efficacy of function-based interventions for students with learning disabilities who exhibit escape-maintained problem behaviors. Preliminary results from a single-case experiment. *Learning Disability Quarterly, 26,* 15-25.

Byrkjedal, I. (2015). *PALS. Skisse til barnehagemodell.* Intern rapport. Oslo: Norsk senter for studier av problematferd og innovative praksis.

C

Caldarella, P., Shatzer, R. H., Gray, K. M., Young, K. R., & Young, E. L. (2011). The effects of school-wide positive behavior support on middle school climate and student outcomes. *RMLE Online, 35,* 1-4.

Card, N. A., Stucky, B. D., Sawalani, G. M., & Little, T. D. (2008). Direct and indirect aggression during childhood and adolescence: A meta-analytic review of gender differences, intercorrelations, and relations to maladjustment. *Child Development, 79,* 1185-1229.

Carr, E. G., Dunlap, G., Horner, R. H., Koegel, R. L., Turnbull, A. P., Sailor, W., Anderson, J. L., Albin, R. W., Koegel, R. W., & Fox, L. (2002). Positive Behavior Support: Evolution of an applied science. *Journal of Positive Behavior Interventions, 4*, 4-20.

Carr, E. G., Horner, R. H., Turnbull, A. P., Marquis, J., Magito-McLaughlin, D., McAtee, M. L., Smith, C. E., Anderson-Ryan, K., Ruef, M. B., & Doolabh, A. (1999). *Positive behavior support for people with developmental disabilities: A research synthesis*. Washington, DC: American Association on Mental Retardation.

Catalano, R., Berglund, M. L., Ryan, G. A. M., Lonczak, H. S, & Hawkins, J. D. (2002). Positive youth development in the United States: Research findings on evaluations of positive youth development programs. *Prevention and Treatment, 5, Article 15,* 1–111.

Charlton, T., & David, D. (eds.) (1993). *Managing misbehavior in schools* (2nd ed.). London: Routledge.

Cicchetti, D., & Rogosh, F. A. (1996). Equifinality and multifinality in developmental psychopathology. *Development and Psychopathology, 8*, 597-600.

Chitiyo, M., May, M. E., & Chitiyo, G. (2012). An assessment of the evidence-base for School-Wide Positive Behavior Support. *Education and Treatment of Children, 35,* 1-24.

Christensen, D. A., Midtbø, T., Helgøy, I., & Homme, A. (2013). Har satsingen Bedre læringsmiljø hatt betydning for elevenes oppfatning av læringsmiljøet? Oslo: Rokkansenteret.

Clark, R. E., & Sugrue, B. M. (1991). *Media in teaching. Handbook of research on teaching*. Media in teaching.

Clausen, M. (2002). *Unterrichtsqualität: Eine frage der perspektive?* Münster: Waxman.

Cochran, W. (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics, 2,* 295-313.

Cochran, W. (1983). *Planning and analysis of observational studies*. New York: Wiley.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciencies.* London: Routledge.

Colvin, G., Kame'enui, E. J., & Sugai, G. (1993). School-wide and classroom management.

Reconceptualizing the integration and management of students with behavior problems in general education. *Education and Treatment of Children, 16*, 361-381.

Colvin, G., Sugai, G., Good, R., & Lee, Y. (1997).Effect of active supervision and precorrection on transition behaviors of elementary students. *School Psychology Quarterly, 12*, 361-381.

Conduct Problems Prevention Research Group (CPPRG, 1999). Initial impact of the fast track prevention trial for conduct problems: II. Classroom effects. *Journal of Consulting and Clinical Psychology, 67*, 648–657.

Cook, T. D. (2003). Why have educational evaluators chosen not to do randomized experiments? *Annals of the American Academy of Political and Social Science, 589*, 114-149.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Chicago, IL: Rand-McNally.

Cook, C. R., Frye, M., Slemrod, T. Lyon, A. R., Renshaw, T. L., & Zhang, Y. (2015). An integrated approach to universal prevention: Independent and combined effects of PBIS and SEL on youths' mental health. *School Psychology Quarterly, 30*, 166-183.

Cook, T. D., Shadish, W. R., & Wong, V. C. (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of policy analysis and management, 27*, 724-750.

Cook, T. D., & Steiner, P. M. (2010). Case matching and the reduction of selection bias in quasi-experiments: The relative importance of pretest measures of outcome, of unreliable measurement, and of mode of data analysis. *Psychological Methods, 1*, 56-68.

Cooper, J. O., Heron, T. E., & Heward, W. L. (1987). *Applied behavior analysis*. Upper Sadle River, NJ: Merrill Publishing.

Creemers, B. P., & Scheerens, J. (1994). Developments in the educational effectiveness research programme. *International Journal of Educational Research, 21*, 125-140.

Cronbach L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika 16*, 297-334.

Critchfield, T. S. (2015). PBIS may not qualify as classical applied behavior analysis. So what? *Behavior Analysis in Practice, 8*, 99–100.

Curby, T. W., Stuhlman, M., Grimm, K., Mashburn, A., Chomat-Mooney, L., Downer, J.,….Pianta, R. C. (2011). Within-day variability in the quality of classroom interactions during third and fifth grade: Implications for childrens' experiences and conducting classroom observations. *The Elementary School Journal, 112*, 16-37.

D

Dagbladet (2006). *Sender ut feil signaler. Professor han ingen tro på belønningssystemet*. 10.07.06

Dagbladet (2015). *Du ser det ikke før du tror det*. 29.04.15

Dalen, M. (2006). *"Så langt det er mulig og faglig forsvarlig …". Inkludering av elever med spesielle behov i grunnskolen.* Oslo: Gyldendal Akademisk.

Davis, H. A. (2003). Conceptualizing the role and influence of student-teacher relationships on

childrens' social and cognitive development. *Educational Psychologist, 38,* 207-234.

Deater-Dekard, K., Dodge, K. A., Bates, J. E., & Pettit, G. S. (1998). Multiple risk factors in the development of externalizing behavior problems: Group and individual differences. *Development and Psychopathology, 10,* 469-493.

Deeks, J. J., Dinnes, J., D'amico, R., Sowden, A. J., Sakarovitch, C., Song, F., ... & Altman, D. G. (2003). Evaluating non-randomised intervention studies. *Health technology assessment (Winchester, England), 7,* iii-x.

De Jong, R., & Westerhof, K. J. (2001). The quality of student ratings of teacher behavior. *Learning Environment Research, 4,* 51-85.

De Los Reyes, A. (2011). Introduction to the special section: More than measurement error: Discovering meaning behind informant discrepancies in clinical assessments of children and adolescents. *Journal of Clinical Child & Adolescent Psychology, 40,* 1-9.

De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A clinical review, theoretical framework, and recommendations for further study. *Psychological Bulletin, 131,* 483-509.

De Pry, R. L., & Sugai, G. (2002). The effect of active supervision and pre-correction on minor behavioral incidents in sixth grade general education classroom. *Journal of Behavioral Education, 11,* 255-267.

DES (Department for Education) (2010). Social and emotional aspects of learning. http://nationalstrategies.standards.dcsf.gov.uk/inclusion/behaviourattendanceandseal

Desimone, L., Smith, T., & Frisvold, D. (2009). Survey measures of classroom instruction: Comparing student and teacher reports. *Educational Policy, 24,* 267-329.

Diekstra, R. F. & Gravesteijn, C. (2008). Effectiveness of school-based social and emotional education programmes worldwide – part one. A review of meta-analytic literature reviews. In C. Clouder (Ed.). *Social and emotional education: an international analysis* (pp. 254-284). Santander, Spain: Fundación Marcelino Botín.

Diekstra, R., Sklad, M., Gravesteijn, C., Ben, J. , & de Ritter, M. (2008). Effectiveness of school-based social and emotional education programmes worldwide - part two. Teaching social and emotional skills worldwide. A meta-analytic review of effectiveness. In C. Clouder (Ed.). *Social and emotional education: an international analysis* (pp. 285–312). Santander, Spain: Fundación Marcelino Botín.

Dirks, M. A., Boyle, M. H., & Georgiades, K. (2011). Psychological symptoms in youth and later socioeconomic functioning: Do associations vary by informant? *Journal of Clinical Child & Adolescent Psychology, 40,* 10-22.

Dix, K. L., Slee, P. T., Lawson, M. J., & Keeves, J. P. (2012). Implementation quality of whole-school mental health promotion and students' academic performance. *Child and Adolescent Mental Health, 17,* 41-51.

Dodge, K. A. (1980). Social cognition and children's aggressive behavior. *Child Development, 51,* 162-170.

Domitrovich, C. E., Bradshaw, C. P., Poduska, J. M., Haagwood, K., Buckley, J., A., Olin, S., et. al. (2008). Maximizing the implementation quailty of evidence-based preventive interventions in

schools: A conseptual framework. *Advances in School Mental Health Promotion, 1*, 6–28.

Domotrovich, C., D., Bradshaw, C., Greenberg, M. T., Embry, D., Poduska, J. M., & Ialongo, N. S. (2010). Integrated models of school-based prevention. Logic and theory. *Psychology in the Schools,* 47:1 71-88.

DuBois, D. L., Holloway, B. E., Valentine, J. C., & Cooper, H. (2002). Effectiveness of mentoring programs for youth: A meta-analytic review. *American Journal of Community Psychology, 30,* 157-198.

Dubow, E. F., Husemann, L. R., Boxer, P., & Smith, C. (2016). Childhood and adloescent risk and protective factors for violence in adulthood. *Journal of Criminal Justice, 45*, 26-31.

Dunlap, G. (2006). The applied behavior analytic heritage of PBS: A dynamic model of action-oriented research. *Journal of Positive Behavior Interventions, 8*(1), 58–60.

Dunlap, G., Strain, P. S., Fox, L., Carta, J. J., Conroy, M., Smith, B. J., et al. (2006). Prevention and intervention with young children´s challenging behavior. Perspectives regarding current knowledge. *Behavioral Disorders, 32*, 29-45.

Durlak, J., & Wells, A. (1997). Primary prevention mental health programs for children and adolescents: a meta-analytic review. *American Journal of Community Psychology, 25*, 115–152.

Durlak, J. A., & Weissberg, R. P. (2007). *The impact of after-school programs that promote personal and social skills*. Chicago, IL: Collaborative for Academic, Social, and Emotional Learning.
Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Schellinger, K. B., & Taylor, R. D. (2011). The impact of enhancing students' social and emotional learning. A meta-analysis of school-based universal interventions. *Child Development, 82, 1,* 405-432.

Dyssegaard, C.B., Larsen, M.S., & Tiftikci, N. (2013). *Effekt og pædagogisk indsats ved inclusion av børn med særlige behov i grundskolen.* Rapport 13. København: Dansk Clearinghouse for Uddannelsesforskning IUP, Aarhus universitet.

E
Eriksen, I. M., Hegna, K., Bakken, A., & Lyng, S. T. (2014). *Felles fokus. En studie av skolemiljøprogram i norsk skole*. Rapport 15/14. Oslo: Velferdsforskningsinstituttet NOVA.

Eklund, H. (2000). Vart är pedagogikforskningen på väg? *Pedagogisk forskning i Sverige, 5*, 131-150.
Ertesvåg, S. K., & Vaaland, G. S. (2007). Prevention and reduction of behavioural problems in school: An evaluation of the Respect program. *Educational Psychology*, *27*, 713-736.

EU (2007). *Towards more knowledge-based policy and practice in education and training*. Commission staff working document, SEC (2007) 1098. Brussel: European Commission, Directorate-General for education and culture.

EVA (2011). *Indsatser for inklusion i folkeskolen.* Rapport. København: Danmarks Evalueringsinstitut.

Evans, I. M., Harvey, S. T., Buckley, L., & Yan, E. (2009) Differentiating classroom climate concepts: Academic, management, and emotional environments, *Kōtuitui: New Zealand Journal of Social Sciences Online, 4*, 131-146.

Eyberg, S. H., Nelson, M. M., & Boggs, S. R. (2008). Evidence-based psychosocial treatments for children and adolescents with disruptive behavior. *Journal of Clinical Child and Adolescent Psychology, 37*, 215-237.

F

Fairchild, G., Van Goozen, S., Stollery, S., Goodyer, I (2008). Fear conditioning and affective modulation of the startle reflex in male adolescents with early-onset or adolescence-onset conduct disorder and healthy control subjects. *Biological Psychiatry, 63,* 279–285.

Farrington, D. P., Gottfredson, D. C., Sherman, L. W., & Welsh, B. C. (2002). The Maryland Scientific Methods Scale. In L. W. Sherman, D. P., Farrington, B. C. Welsh & D. L. MacKenzie (Eds.), *Evidence-Based Crime Prevention* (pp. 13-21). New York: Routledge.

Feingold, A. (2013). A regression framework for effect size assessments in longitudinal modeling of group differences. *Review of General Psychology, 1,* 111-121.

Fenning, P., Theodos, J., Benner, C., & Bohanonon-Edmonson, H. (2004). Integrating proactive discipline practices into codes of conduct. *Journal of School Violence, 3,* 45-61.

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice, 40,* 532-538

Ferguson, R. F. (2012). Can student surveys measure teaching quality? *Phi Delta Kappan, 94,* 24-28.

Fergusson, D. M., Horwood, J., & Ridder, E. M. (2007). Conduct and attentional problems in childhood and adolescence and later substance use, abuse and dependence: Results of a 25-year longitudinal study. *Drug and Alcohol Dependence, 88,* 14-26.

Field, A. (2009). *Discovering statistics using SPSS (3rd ed.).* London: Sage Publication Ltd.

Fisher, D. L., & Fraser, B. J. (1981). Validity and use of My Class Inventory. *Science Education, 65,* 145-156.

Fixsen, D. L., Naoom, S. F., Blasé, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: a synthesis of the literature*. http://www.fpg.unc.edu/~nirn/resources/publications/Monograph/

Flake, J. K., & Petway, K. T. (2019). Methodologies for investigating and interpreting student-teacher rating incongruence in noncognitive assessment. *Educational Measurement: Issues and Practices, 38,* 63-77.

Flannery, K. B., Fenning, P., Kato, M. M., & McIntosch, K. (2014). Effects of school-wide positive behavioral interventions and supports and fidelity of implementation on problem behavior in high schools. *School Psychology Quarterly, 29,* 111-124.

Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., ... & Ji, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention science, 6,* 151-175.

Forgatch, M. S., Patterson, G. R., & DeGarmo, D. S. (2005). Evaluating fidelity: predictive validity for a measure of competent adherence to the Oregon model of Parent Management Training. *BehaviorTherapy, 36,* 3–13.

Forns, M., Abad, J., & Kirchner, T. (2014). Internalizing and externalizing problems. *Encyclopedida of Adolescence,* pp 1464-1469.

Frazer, M. W., & Galinsky, M. J. (1997). Toward a resilience-based model of practice. In M. W. Fraser (Ed.), *Risk and resilience in childhood: An ecological perspective* (pp. 265-275). Washington;

DC: NASW Press.

Fraser, B. J., & O`Brien, P. (1985). Student and teacher perceptions of the environment of elementary school classrooms. *The Elementary School Journal, 85,* 567-580.

Freeman, J., Simonsen, B., McCoach, D. B., Sugai, G., Lombardi, A., & Horner, R. (2015). An analysis of the relationship between implementation of school-wide positive behavior interventions and supports and high school dropout rates. *The High School Journal, 98,* 290-315.

Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., et al. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science, 6,* 151–175.

Fuchs, D., & Deshler, D. D. (2007). What we need to know about Responsiveness to Intervention (and shouldn´t be afraid of). *Learning Disabilities, 22,* 129-136.

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

G

Gage, N. A., Larson, A., Sugai, G., & Chafouleas, S. M. (2016). Student perceptions of school climate as predictors of office discipline referrals. *American Educational Research Journal, 55,* 492-515.

Gage, N. A, Whitford, D. K., & Katsiyannis, A. (2018). A review of Schoolwide Positive Behavior Interventions and Supports as a framework for reducing disciplinary exclusions. *The Journal of Special Educations, 52,* 142-151.

Garmezy, N. (1993). Developmental psychopathology. Some historical and current perspectives. In D. Magnusson & P. Casaer (Eds.), *Longitudinal research on individual development. Present status and future perspectives* (pp. 95-126). Cambridge: Cambridge University Press.

Garmezy, N., & Rutter, M. (1993). *Stress, coping, and development in children*. New York: McGraw-Hill.

Garrard, W., & Lipsey, M. (2007). Conflict resolution education and anti-social behavior in US schools. A meta-analysis. *Conflict Resolution Quarterly, 25,* 9–37.

Gershenson, S. (2016). Linking teacher quality, student attendance, and student achievement. *Education Finance and Policy, 11,* 125-149.

Gibson, S., & Dembo, M. H. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology, 76,* 569- 582.

Given, L. M. (2007). Descriptive research. In. N. J. Salkind, & K. Rasmussen (Eds.), *Encyclopedia of measurement and statistics (*pp. 251-254). Thousand Oaks, CA: Sage.

Givvin, K. B., & Santagata, R. (2011). Toward a common language for discussing the features of effective professional development: The case of a US mathematics program. *Professional Development in Education, 37,* 439–451.

Gjengangeren (2014). *Mener PALS er bortkastet*. 02.01.14

Goddard, R. D. (2002). A theoretical and empirical analysis of the measurement of collective efficacy:

The development of a short form. *Educational and Psychological Measurement, 62,* 97-110.

Goddard, R. D. & Goddard, Y. L. (2001). A multilevel analysis of the relationship between teacher and collective efficacy in urban schools. *Teaching and Teacher Education, 17,* 807-818.

Goddard, R. D., Hoy, W. K., & Woolfolk Hoy, A. (2000). Collective efficacy: Its meaning, measure, and impact on student achievement. *American Educational Research Journal, 37,* 479-507.

Goddard, R. D., Hoy, W. K., & Woolfolk Hoy, A. (2004). Collective efficacy beliefs: Theoretical developments, empirical evidence, and future directions. *Educational Researcher, 33,* 470-507.

Gottfredson, G. D., Gottfredson, D. C., Payne, A., & Gottfredson, N. C. (2005). School climate predictors of school disorder. Results from national delinquency prevention in school. *Journal of Research in Crime and Delinquency, 42,* 421-444.

Goldstein, S. E., Young, A., & Boyd, C. (2008). Relational aggression at school. Associations with school safety and social climate. *Journal of Youth & Adolescence, 37,* 641-654.

Gordon, R. S. (1983). An operational classification of disease prevention. *Public Health Reports, 98,* 107–109.

Grazziano, A. M., & Raulin, M. L. (2007). *Research methods: A process of inquiry (6th ed.).* Boston: Allyn and Bacon.

Gravetter, F. J., & Forzano, L. A. B. (2011). Research methods for the behavioral sciences (4th ed.). Belmont CA: Wadsworth,

Gregory, A., Cornell, D., Fan, X., Sheras, P., Shih, T., & Huang, F. (2010). Authoritative school discipline. High school practices associated with lower student bullying and victimization. *Journal of Educational Psychology, 102,* 483-496.

Gresham, F. M., & Elliott, S. N. (1990). *Social skills rating system—Manual*. Circle Pines: American Guidance Service.

Gresham, F. M., Elliott, S. N., Cook, C. R., Vance, M. J., & Kettler, R. (2010). Cross-informant agreement for ratings for social skill and problem behavior ratings: an investigation of the Social Skills Improvement System-Rating Scales. *Psychological Assessment, 22,* 157-166.

Gresham, F. M., Elliott, S. N., Metallo, S., Byrd, S., Wilson, E., & Cassidy, K. (2018). Cross-informant agreement of children's social-emotional skills: An investigation of ratings by teachers, parents, and students from a nationally representative sample. *Psychology in the Schools, 55,* 208–223.

Grey, J., & Sime, N. (1989). Findings from the national survey of teachers in England and Wales. In Elton . *Discipline in Schools*. Report of the committee of enquiry chaired by Lord Elton. Department of Education and Science and the Welsh office (pp. 222-292). London: Her Majesty's Stationary Office.

Griffin, B. W. (2004). Grading leniency, grade discrepancy, and student ratings of instruction. *Contemporary Educational Psychology, 29,* 410-425.

Grissom, R. J., & Kim, J. J. (2005). *Effect sizes for research. A broad practical approach.* Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Grossman, D. C., Neckerman, H. J., Koepsell, T. D., Liu, P., Asher, K. N., Beland, K., Frey, K., & Rivera, F. P. (1997). Effectiveness of a violence prevention curriculum among children in elementary

school. A randomized controlled trial. *The Journal of the American Medical Association, 27*, 1605-1612.

Guo, P., Higgins-D'Alessandro (2011, October). *The place for teachers' views of teaching in promoting positive school culture and student prosocial and academic outcomes*. Paper presented at the Association for Moral Education annual conference, Nanjing, China.

Gustafson, J. E., Allodi, M. W., Alin,B. Å., Eriksson, C., Eriksson, L., Fischbein, S. Granlund, M., Gustafsson, P., Ljungdahl, S., Ogden, T., & Persson, R. S. (2010). *School learning and mental health. A systematic review*. Report. Stockholm: The Health Committee, The Royal Swedish Academy of Science.

Gutman, L.M., Joshi, H., Parsonage, M., & Schoon, I. (2018). Gender-specific trajectories of conduct problems from ages 3 to 11. *Journal of Abnormal Child Psychology, 46*, 1467–1480.

H

Hagopian, L. P., Hardesty, S. L., & Gregory, M. (2017). *Overview and summary of scientific support for applied behavior analysis.* The Kennedy Krieger Institute and Johns Hopkins University School of Medicine. Retrieved from https://www.kennedykrieger.org/sites/default/files/patient-care-files/aba-scientific-support-9-2015.pdf

Hahn, R. (2007) Task force on community preventive services: effectiveness of universal school-based programs to prevent violent and aggressive behavior: a systematic review. *American Journal of Preventive Medicine, 33*, 114–129.

Hahn, R., Fuqua-Whitley, D., Wethington, H., Lowy, J., Crosby, A., Fullilove M., ….Dahlberg L. (2007). Effectiveness of universal school-based programs to prevent violent and aggressive behaviour: A systematic review. *American Journal of Preventive Medicine, 33* (Suppl. 2), 114–129.

Hamre, B. K., & Pianta, R. C. (2001). Early teacher child relationships and the trajectory of childrens' school outcomes through eighth grade. *Child Development, 72*, 949-967.

Hartley, A., Zakriski, A. L., & Wright, J. C. (2011). Probing the depths of informant discrepancies: Contextual influences on divergence and convergence. *Journal of Clinical Child & Adolescent Psychology, 40*, 54-66.

Hattie, J. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. New York & London: Routledge

Hawken, L., & Horner, R. H. (2003). Evaluation of a targeted group intervention within a schoolwide system of behaviour support. *Journal of Behavioural Education, 3*, 225–240.

Hawken, L. S., MacLeod, K. S., & Rawlings, L. (2007). Effects of the Behavior Education Program (BEP) on office discipline referrals of elementary school students. *Journal of Positive Behavior Interventions, 9*, 94-101.

Hawkins, J. D., Catalano, R. F., Kosterman, R., Abbott, R., & Hill, K. G. (1999). Preventing adolescent health-risk behaviors by strengthening protection during childhood. *Archives of Pediatrics and Adolescent Medicine, 153*, 226-234.

Hawkins, J. D., Catalano, R. F., & Miller, J. Y. (1992). Risk and protective factors for alcohol and other drug problems in adolescence and early adulthood: Implications for substance abuse prevention. *Psychological Bulletin, 112*, 64-105.

Healthy Schools (2011). http://home.healthyschools.gov.uk/

Heck, R. H., Thomas, S. L., & Tabata, L. N. (2010). *Multilevel and longitudinal modeling with SPSS*. New York/London: Routledge.

Hedges, L. V. & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis, 29*, 60-87.

Hedges, L. V., & Rhoads, C. (2010). *Statistical power analysis in education research*. NCSER 2010-3006. Washington, DC: National Center for Special Education Research, Institute for Education Sciences, US Department of Education.

Heinsman, D. T., & Shadish, W. R. (1996). Assignment methods in experimentation: When do nonrandomized experiments approximate answers from randomized experiments? *Psychological Methods, 1*, 154-169.

Helgøy, I., & Homme, A. (2015). *Økt innsats for læringsmiljøet. Evaluering av den nasjonale satsingen Bedre læringsmiljø.* Delrapport 5. Oslo: Rokkansenteret.

Helsinki D. (2010). Ethical principles for medical research involving human subjects. http://www.etikkom.no/retningslinjer/helsinkideklarasjonen.

Heiervang, E., Stormark, K. M., Lundervold, A. J., Heiman, M., Goodman, R., Posserud, M. B., . . . Gilleberg, C. (2007). Psychiatric disorders in Norwegian 8-to-10-year-olds: An epidemiological survey of prevalence, risk factors, and service use. *Journal of the American Academy of Child and Adolescent Psychiatry 46*, 438-447.

Helgeland, I. M. (2008). *Forebyggende arbeid i skolen. Om barn med sosiale og emosjonelle vansker*. Oslo: Kommuneforlaget.

Heward, W. L. (2006). *Exceptional children. An introduction to special education* (8th ed.). Upper Saddle River, NJ: Pearson.

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect size research. *Child Development, 2*, 172.177.

Hill, H. C., & Ericson, A. (2019). Using implementation fidelity to aid interpreting program impacts: A brief review. *Educational Researcher, 48*, 590-598.

Hirschi, T. (1998). *Social bond theory*. *Criminological theory: Past to present.* Los Angeles: Roxbury

Horner, R.H., Koegel, R. L., Carr, E. G., Sailor, W., Anderson, J., Albin, R. W., & O'Neill, R. E. (1990). Toward a technology of 'nonaversive' behavioral support. *Research & Practice for Persons with Severe Disabilities, 5*, 3-10.

Horner, R. H., Sugai, G, & Lewis, T. (2015). *Is School-Wide Positive Behavior Support an evidence-based practice?* https://www.pbis.com

Horner, R. F., & Sugai, G. (2015). School-wide PBIS: an example of applied behavior analysis. Implemented at a scale of social importance. *Behavior Analysis in Practice, 8*, 80–85.

Horner, R. H., Sugai, G. & Anderson, C. M. (2010). Examining the evidence base for School-Wide Positive Behavior Support. *Focus on Exceptional Children, 42*, 1-14.

Horner, R. H., Sugai, G. & Lewis, T. (2015). *Is School-Wide Positive Behavior Support an evidence-*

based practice? www.pbis.com

Horner, R. H., Sugai, G., Smolkowski, K., Eber, L., Nakasato, J., Todd, A. W., & Esperanza, J. (2009). A randomized, wail-list controlled effectiveness trial assessing school-wide positive behavior support in elementary schools. *Journal of Positive Behavior Interventions, 11*, 133-144.

Horner, R. H., Todd, A. W., Lewis-Palmer, T., Irvin, L. K., Sugai, G., & Boland, J. B. (2004). The School-Wide Evaluation Tool (SET). A research instrument for assessing School-Wide Positive Behavior Support. *Journal of Positive Behavior Interventions, 6,* 3-12.

Horowitz, J. L., & Garber, J. (2006). The prevention of depressive symptoms in children and adolescents: a meta-analytic review. *Journal of Consulting and Clinical Psychology, 74,* 401–415.

Horton, N. J., Roberts, K., Ryan, L., Suglia, S. F., & Wright, R. J. (2008). A maximum likelihood latent variable regression model for multiple informants. *Statistics in medicine, 27,* 4992-5004.

Houchens, G. W., Zang, J., Davis, K., Niu, C., Chon, K. H., & Miller, S. (2017). The impact of positive behavior interventions and supports on teachers' perceptions of teaching conditions and student achievement. *Journal of Positive Behavior Interventions, 19,* 168-179.

Houts, R. M., Caspi, A., Pianta, R. C., Arseneault, L., & Moffitt, T. E. (2010). The challenging pupil in the classroom. The effect of the child on the teacher. *Psychological Science, 2,* 1802-1810.

Hox, J. J. (2010). *Multilevel analysis. Techniques and applications (2nd ed.).* New York/Hove: Routledge.

Hsieh, P., Acee, T., Chung, W., Hsieh, Y., Kim, H., Thomas, G. D., & Robinson, D. H. (2005). Is educational intervention research on the decline? *Journal of Educational Psychology, 97,* 523-529.

Hukkelberg, S., & Sørlie, M-A. (in review). *A multilevel confirmatory factor analysis of the Short Collective Teacher Efficacy scale*. Journal of Psychoeducational Assessment.

Humphrey, N., Lendrum, A., Barlow, A., Wigelsworth, M., & Squires, G. (2013). Achievement for all: Improving psychosocial outcomes for students with special educational needs and disabilities. *Research in Developmental Disabilities, 4,* 1210-1225.

Humphrey, N., Lendrum, A., & Wigelsworth, M. (2013). Making the most out of school-based prevention: Lessons from the social and emotional aspects of learning (SEAL) programme. *Emotional and Behavioural Difficulties, 18,* 248-260.

I

Ialongo, C. (2016). Understanding the effect size and its measures. *Biochemia Medica, 26,* 150-163.

Ialongo, N. S., Rogosch, F. A., Cicchetti, D., Toth, S. L., Buckley, J., Petras, H.,… et al.. (2006). A developmental psychopathology approach to the prevention of mental health disorders. In D. Cicchetti (Ed.), *Developmental psychopathology* (2nd ed., pp. 968-1018). New York: Wiley.

Idsoe, T., Dyregrov, A., & Idsoe, E. C. (2012). Bullying and PTSD symptoms. *Journal of Abnormal Child Psychology, 40,* 901-911.

Ingemarson, M., Rosendahl, I., Bodin, M., & Birgegård, A. (2019). Teacher's use of praise, clarity of school rules and classroom climate: comparing classroom compositions in terms of disruptive students. *Social psychology of education, 23,* 217–232

Iovannone, R., Greenbaum, P.E., Wang, W., Kincaid, D., Dunlap, G., & Strain, P. (2009). Randomized controlled trial of the prevent-teach-reinforce (PTR) tertiary intervention for students with problem behaviors. *Journal of Emotional and Behavioral Disorders, 17*, 213-225.

J

Jackson, C. K. (2018). What does test scores miss? The importance of teacher effects on non-test score outcomes. *Journal of Political Economy, 126*, 2072-2107.

Jacob, R. T., Doolittle, F., Kemple, J., & Somers, M.-A. (2019). A framework for learning from null results. *Educational Researcher, 48*, 580–589.

Jahnsen, H., Nergaard, S. E., & Flaatten, S. V. (2006). *I randsonen. Forekomst og organisering av smågruppetiltak for elever på ungdomstrinnet som viser problematferd og lav skolemotivasjon.* Porsgrunn: Lillegården kompetansesenter.

Jarmolowicz, D. P., & Tetreault, A. S. (2015). Positive behavioral interventions and supports as behavior analysis in practice: Why is this even controversial? *Behavior Analysis in Practice, 8*, 86-87.

Jennings, J. L., & DiPrete, T. A. (2010). Teacher effects on social and behavioral skills in early elementary school. *Sociology of Education, 83*, 135-159.

Jessor, R., Turbin, M. S., & Costa, F. M. (1998). Protective factors in adolescent health behavior. *Journal of Personality and Social Psychology, 75*, 788-800.

Jessor, R., Van Den Bos, J., Vanderryn, J., Costa, F. M., & Turbin, M. S. (1995). Protective factors in adolescent problem behavior: Moderator effects and developmental change. *Developmental Psychology, 31*, 923-933.

Johnson, V. A., Kemp, A. H., Heard, R., Lennings, C. J., & Hickie, I. B. (2015). Childhood-versus adolescent-onset antisocial youth with conduct disorder; Psychiatric illness, neuropsychological and psychosocial function. *PLoS One, 10*(4): e0121627. doi: 10.1371/journal.pone.0121627

Jolivette, K., Patterson, D. P., Swoszowski, N. C., McDaniel, S. C., Kennedy, C., & Ennis, R. P. (2014). School-wide positive behavioral interventions and supports in a residential school for students with emotional and behavioral disorders: First years of implementation and maintenance follow-up focus groups. *Residential Treatment for Children & Youth, 1*, 63-79.

Jüni, P., Altman, D. G., & Egger, M. (2001). Assessing the quality of controlled clinical trials. *British Medical Journal, 323*, 42-46.

K

Kauffman, J. M. (1996). Research to practice issues. *Behavioral Disorders, 22*, 55–60.

Keenan, K., & Shaw, D. (1997). Developmental and social influences of young girls' early problem behavior. *Psychological Bulletin, 121*, 95-113.

Kelm, J. L., & McIntosh, K. (2012). Effects of school-wide positive behavior support on teacher self-efficacy. *Psychology in the Schools, 49*, 137-147.

Kelm, J. L., McIntosh, K., & Cooley, S. (2014). Effects of implementing school-wide positive behavioural interventions and supports on problem behaviour and academic achievement in a Canadian elementary school. *Canadian Journal of School Psychology, 195-212*. Doi: 10.1177/082957354540266.

Keogh-Brown, M. R., Bachmann, M. O., Shepstone, L., Hewitt, C., Howe, A., Ramsay, C. R., ... & Elbourne, D. R. (2007). Contamination in trials of educational interventions. Health Technology Assessment [Online], 11(43). http://www.ncchta.org/fullmono/mon1143.pdf

Kern, L., Childs, K. E., Dunlap, G., Clarke, S., &Falk, G. D. (1994). Using assessment-based curricular intervention to improve the classroom behavior of a student with emotional and behavioral challenges. *Journal of Applied Behavior Analysis, 27*, 7-19.

Kern, L., & Manz, P. (2004). A look at current validity issues of School-Wide Behavior Support. *Behavioral Disorders, 30*, 47-59.

Kidsmatter (2009). http://www.kidsmatter.edu.au

Killerby, P., & Dunsmuir, S. (2018). Is implementation of evidence-based interventions in schools related to pupil outcomes? A systematic review. *Educational and child psychology. Special issue.* 108-121.

Kim, J. S. (2019). Making every study count: Learning from replication failure to improve intervention research. *Educational Researcher, 48*, 599–607.

Kimber, B. (2011). *Primary prevention of mental health problems among children and adolescents through social and emotional training in school*. Thesis for doctoral degree. Stockholm: Department of Public Health Sciences, Division of Social Medicine, Karolinska Institutet.

Kincaid, D., & George, H. (2012). *District action planning and problem-solving process for multi-tiered academic and behavior systems*. Presentation at School-wide PBIS Implementers' Forum, Oct.18, 2012. Chicago, IL.

Kjøbli, J., & Sørlie, M. A. (2008). School outcomes of a community-wide intervention model aimed at preventing problem behavior. *Scandinavian Journal of Psychology*, *49*, 365-375.

Koedel, C. (2008). Teacher quality and dropout outcomes in a large, urban school district. *Journal of Urban Economics, 64*, 560-572.

Kopershoek, H., Harms, T., de Boer, H., van Kuijk, M., & Doolaard, S. (2016). A meta-analysis of the effects of classroom management strategies and classroom management programs of students' academic, behavioral, emotional, and motivational outcomes. *Review of Educational Research, 86*, 643-680.

Kownacki, R. J., & Shadish, W. R. (1999). Does Alcoholics Anonymous work? The results from a meta-analysis of controlled experiments. *Substance Use & Misuse, 34*, 1897-1916.

Kraemer, H. C., Measelle, J. R., Ablow, J. C., Essex, M. J., Boyce, W. T., & Kupfer, D. J. (2003). A new approach to integrating data from multiple informants in psychiatric assessment and research: Mixing and matching contexts and perspectives. *American journal of psychiatry*, *160*, 1566-1577.

Kraft, M. (2019). *Interpreting effect sizes of education interventions*. (EdWorkingPaper: 19-10). Annenberg Institute, Brown University. http://www.edworkingpapers.com/ai19-10

Kraft, M. A., & Grace, S. (2016). *Teaching tomorrow's economy? Teacher effects on complex cognitive skills and social-emotional competencies* (Working paper). Providence, RI: Brown University.

Kroll, L., Rothwell, J., Bradley, D., Shah, P., Baiey, S., & Harrington, R. (2002). Mental health needs of boys in secure care for serious or persistent offending: a prospective, longitudinal study. *Lancet,*

359, 1975–1979. doi: 10.1016/S0140-6736(02)08829-3

Kjærnsli, M., Lie, S., Olsen, R.V., Roe, A., & Turmo, A. (2004). *Rett spor eller ville veier? Norske elevers prestasjoner i matematikk, naturfag og lesing i PISA 2003*. Oslo: Universitetsforlaget.

Kjærnsli, M., & Olsen, R. V. (red.) (2013). *Fortsatt en vei å gå. Norske elevers kompetanse i matematikk, naturfag og lesing i PISA 2012*. Oslo: Universitestforlaget.

Kopershoek, H., Harms, T., de Boer, H., van Kuijk, M., & Doolaard, S. (2016). A meta-analysis of the effects of classroom management strategies and classroom management programs on students' academic,
behavioral, emotional, and motivational outcomes. *Review of Educational Research, 86,* 643-680.

Kristoffersen, J. H. G., Obel, C., & Smith, N. (2014). Gender differences in behavioral problems and school outcomes. *Journal of Economic Behavior & Organization, 115*, 75-93.

Kokkersvold, E. (2012). Barn og unge som viser alvorlig problematferd – er vi på rett vei? *Bedre Skole, 4*, 28-31.

KUF/BFD (2000). *Vurdering av program og tiltak for å redusere problematferd og utvikle sosial kompetanse* Innstilling fra faggruppe oppnevnt av Kirke-, utdannings- og forskningsdepartementet og Barne- og familiedepartementet. Oslo: Kirke-, utdannings- og forskningsdepartementet.

Kunter, M., & Baumert, J. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environment Research, 9*, 231-251.

L

Ladd, H. F., & Sorensen, L. C. (2015*). Returns to teacher experience: Student achievement and motivation in middle school* (Working paper, 112). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Lahey, B. B., Van Hulle, C. A., Waldman, I. D., Rodgers, J. L., D'Onofrio, B. M. et al. (2006). Testing descriptive hypotheses regarding sex differences in the development of conduct problems and delinquency. *Journal of Abnormal Child Psychology, 34*, 737-755.

Landsberger, H. A. (1958). *Hawthorne Revisited*. Ithaca Press.

Lane, K. L., & Menizies, H. M. (2003). A school-wide intervention with primary and secondary levels of support for elementary students. Outcomes and considerations. *Education and Treatment of Children, 26,* 431-451.

Lannie, A. L., Codding, R. S., McDougal, J. L., & Meier, S. (2010). The use of change-sensitive measures to assess school-based therapeutic interventions. Linking theory and practice at the tertiary level. *School Psychology Forum, 4*, 1-14.

Lassen, S., Steele, M., & Sailor, W. (2006). The relationship of school-wide positive behavior support to academic achievement in an urban middle school. *Psychology in Schools, 43*, 701-712.

Latham, G. (1992). Interacting with at-risk children. The positive position. *Principal, 72*, 26-30.

Lee, A., & Gagne, N. A. (2020). Updating and expanding systematic reviews and meta-analyses on the effects of school-wide positive behavior interventions and supports. *Psychology in the Schools, 57*, 1-22. https://doi.org/10.1002/pits.22336

Lewis, T. J., Barrett, S., Sugai, G., & Horner, R. H. (2010). *Blueprint for schoolwide positive behavior support training and professional development.* Eugene, OR: National Technical Assistance Center on Positive Behavior Interventions and Support.

Lewis, T. J, Colvin, G. &.Sugai, G. (2000). The effects of pre-correction and active supervision on the recess behavior of elementary students. *Education & Treatment of Children, 23,* 109-121.

Lewis, T., Hudson, S., Richter, M., & Johnson, N. (2004). Scientifically supported practices in EBS. A proposed approach and brief review of current practices. *Behavior Disorders, 29,* 247-259.

Lewis, T. J., Powers, L. J., Kelk, M. J., & Newcomer, L. (2002). Reducing problem behaviors on the playground: An investigation of the application of school-wide positive behavior supports. *Psychology in the Schools, 39,* 181-190.

Lindberg, E., & Ogden, T. (2001). *Elevatferd og læringsmiljø 2000. En oppfølgingsundersøkelse av elevatferd og læringsmiljø i grunnskolen. Rapport*. Oslo: Læringssenteret.

Lipsey, M. W. (2000). Statistical conclusion validity for intervention research. A significant ($p < .05$) problem. In L. Bickman, (Ed.). *Vaidity and social experimentation: Donald Campbell's legacy*. Thousand Oaks, CA: Sage.

Lipsey, M. W., Puzio, K., Yun, C, Herbert, M.A., Steinka-Fry, C., Cole, M. W.,….Busick, M. D. (2012*). Translating the statistical representation of the effects of education interventions into more readily interpretable forms*. Washington, DC: U. S. Department of Education, Institute of Educational Science, National Center for Special Education Research. *Journal of Positive Behavior Interventions, 20,* 239-250.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: confirmation from meta-analysis. *American psychologist, 48,* 1181-1209.

Little, R. J. A., & Rubin, D. (2002). *Statistical analysis with missing data (2nd ed.)*. Hoboken, NJ: John Wiley & Sons, Inc.

Lleras (2008). Hostile school climates. Explaining differential risk of student exposure to disruptive learning environments in high school. *Journal of School Violence, 7,* 105-135.

Loeber, R., & Farrington, D. P. (2000) Young children who commit crime: Epidemiology, developmental origins, risk factors, early interventions, and policy implications. *Development and Psychopathology, **12**,* 737-762.

Loeber, R., Farrington, D. P., Stouthamer-Loeber, M., Moffitt, T. E., & Caspi, A. (1998). The development of male offending: Key findings from the first decade of the Pittsburgh Youth Study. *Studies on Crime & Crime Prevention, 7,* 141-171.

Lohrmann-O'Rourke, S., Kostner, T., Sabatine, K., Smith, D., Horvath, B., & Llewellyn, G. (2000). School-wide application of PBS in the Bangor area school district. *Journal of Positive Behavior Interventions, 2,* 238-240.

Loukas, A. K. (2015). PBIS is (not) behavior analysis: a response to Horner and Sugai (2015). *Behavior Analysis in Practice, 8,* 95–98.

Luiselli, J. K, Putnam, R. F, Handler, M. W, & Feinberg, A. B. (2005). Whole-school positive behaviour support: Effects on student discipline problems and academic performance. *Educational Psychology, 25,* 183-198.

M

Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology, 1*, 86-92.

March, H. W., & Roche, L. A. (2000). *The use of studnets' evaluations of university teaching to improve*
teaching effectiveness. Canberra, Australian Capital Territory, Australia: Department of Employment, Education and Training.

Markussen, E. (2002). *Særskilt tilrettelagt opplæring i videregående – hjelper det? Om segregering, inkludering og kompetanseoppnåelse i det første R-94 kullet.* Rapport nr. 341. Oslo: Fafo. .

Markussen, E., Strømstad, M., Carlsson, T. C., Haustätter, R., & Nordahl, T. (2007). *Inkluderende spesialundervisning? Om utfordringer innenfor spesialundervisningen i 2007.* Rapport nr. 1 fra prosjektet: Gjennomgang av Spesialundervisning, Evaluering av Kunnskapsløftet. Rapport 19:2007. Oslo: NIFU STEP.

Marzano, R. J., Marzano, J. S., & Pickering, D. J. (2003). *Classroom management that works. Research-based strategies for every teacher*. 1st ed. Alexandria, VA: Association for Supervision and Curriculum Development.

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D.,….Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development, 79,* 732-749.

Masten, A. S. (2006). Promoting resilience in development: A general framework for systems of care. *Promoting resilience in child welfare, 2*.

Matajasko, J. L., Vivi-Kantor, A. M., Massetti, G. M., Holland, K. M., Holt, M. K., & Cruz, J. D. (2012). A systematic meta-review of evaluations of youth violence prevention programs: Common and divergent findings from 25 years of meta-analyses and systematic reviews. *Aggression and Violent Behavior, 17,* 540-552.

Mattison, E., & Aber, M. S. (2007). Closing achievement gap. The association of racial climate with achievement and behavioral outcomes. *American Journal of Community Psychology, 40,* 1-12.

Maxwell, S. E., Kelley, K., & Rausch, J. R. (2008). Sample size planning for statistical power and accuracy in parameter estimation. *Annual Review of Psychology, 59,* 537-563.

Mayer, G. (1995. Preventing antisocial behavior in the schools. *Journal of Applied Behavior Analysis, 28,* 467-478.

Maysiak, R. (1997). Exploring the theory and paradigm base for wraparound. *Journal of Child and Family Studies, 6, 399-408.*

McCart, M. R, Ogden, T., & Henggeler, S. W. (2014). Youth who have broken the law. In G. B. Melton, Ben-Arieh, A., Cashmore, J., Goodman, G. S., & Worley, N. K. (Eds.). *The SAGE handbook of child research*. Los Angeles: Sage.

McCartney, K., & Rosenthal, R. (2000). Effect size, practical importance, and social policy for children. *Child Development, 71,* 173-180.

McCurdy, B. L., Mannella, M. C., & Eldridge, N. (2003). Positive behavior support in urban schools. Can we prevent the escalation of antisocial behavior? *Journal of Positive Behavior Interventions, 5,*

158-170.

McIntosh, K., Bennett, J. L., & Price, K. (2011) Evaluation of social and academic effects of School-Wide Positive Behavior Support in a Canadian school district. *Exceptionality Education International, 21*, 46-60.

McNeely, C. A, Nonnemaker, J. M., & Blum, R. W. (2002). Promoting student connectedness to school. Evidence from the national longitudinal study of adolescent health. *Journal of School Health, 72*, 138-146.

Metzler, C. W., Biglan, A., Rusby, J. C., & Sprague, J. R. (2001). Evaluation of a comprehensive behavior management program to improve school-wide positive behavior support. *Education and Treatment of Children, 24,* 448-479.

Midthassel, U. V., & Ertesvåg, S. K. (2008). Schools implementing Zero: The process of implementing an anti-bullying program in six Norwegian compulsory schools. *Journal of Educational Change, 9,* 153-172.

Miller, E. (2002). Brain injury as a contributory factor in offending. In J. Glickson (Ed.). *The neurobiology of criminal behavior* (pp. 134–156). Dordrecht: Kluwer Academic

Mind Matters (2009). www.mindmatters.edu.au

Moffitt, T. E. (1993). Adolescent-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review, 100,* 674-701.

Moos, R. H., & Trickett, E. J. (1974). *The classroom environment scale*. Los Angeles, CA: Consulting Psychology Press.

Morrison, G. M., Robertson, L., Laurie, & Kelly, J. (2002). Protective factors related to antisocial behavior trajectories. *Journal of Clinical Psychology, 58,* 277-290.

Mrazec, P. G., & Haggerty, R. J. (Eds.) (1994). *Reducing risk for mental health disorders. Frontiers for preventive intervention research*. Washington, DC: National Academic Press.

Murray, J., & Farrington, D. P. (2010). Risk factors for conduct disorders and delinquency: Key findings from longitudinal studies. *The Canadian Journal of Psychiatry, 55*, 633-640.

Muscott, H. S., Mann, E. L., & LeBrun, M. R. (2008). Positive behavioral interventions and supports in New Hampshire: Effects of large-scale implementation of schoolwide positive behavior support on student discipline and academic achievement. *Journal of Positive Behavior Interventions, 10,* 90-105.

Myers, D., Simonsen, B. & Sugai, G. (2011). Increasing teachers' use of praise with response to intervention approach. *Education & Treatment of Children, 34,* 35-59.

Mytton, J. A., DiGuiseppi, C., Gough, D., Taylor, R. S., & Logan, S. (2002) School-based secondary prevention programmes for preventing violence. *Cochrane Database of Systematic Reviews 2006, Issue 3, 1–93,* Art. No.: CD004606.

N

Nelson, J. R. (1996). Designing schools to meet the needs of students who exhibit disruptive behavior. *Journal of Emotional and Behavioral Disorders, 4,* 147-161.

Nelson, J.R., Martella, R., & Marchand-Martella, N. (2002). Maximizing student learning: The effects

of a comprehensive school-based program for preventing problem behaviors. *Journal of Emotional and Behavioral Disorders, 10,* 136-148.

Newcomer, L. L., & Lewis, T. J. (2004). Functional behavior assessment. An investigation of assessment reliability and effectiveness of function-based interventions. *Journal of Emotional and Behavioral Disorders, 12*, 168-181.

NFR (Norges Forskningsråd) (2014). *Utdanning og forskning i spesialpedagogikk – veien videre.* Rapport fra Ekspertgruppen for spesialpedagogikk. Oslo: Norges Forskningsråd.

NIFU (2010). «*Hvis noen forteller om mobbing". Utdypende undersøkelse av funn fra Elevundersøkelsen om mobbing, urettferdig behandling og diskriminering*. Rapport 48 v. Lødding, B. & Vibe, N. Oslo: Nordisk institutt for studier av innovasjon, forskning og utdanning.

Noltemeyer, A., Palmer, K., James, A. G., & Wiechman, S. (2019). School-Wide Positive Behavioral Interventions and Supports (SWPBIS): A synthesis of existing research. *International Journal of School & Educational Psychology, 7,* 253-262.

Nordahl, T. (2009): Realisering og resultater av tilpasset opplæring i grunnskolen». I: T. Nordahl og S. Dobson (red.): *Skolen og elevens forutsetninger. Om tilpasset opplæring i pedagogisk praksis og forskning*. Vallset: Opplandske bokforlag.

Nordahl, T., Gravrok, Ø., Knutsmoen, H., Larsen, T. M. B., & Rørnes, K. (red.) (2006). *Forebyggende innsatser i skolen. Rapport fra forskergruppe oppnevnt av Utdannings-direktoratet og Sosial- og helsedirektoratet om problematferd, rusforebyggende arbeid, og læreren som leder og implementeringsstrategier.* Oslo: Utdanningsdirektoratet og Sosial- og helsedirektoratet.

Nordahl, T., Persson, B., Dyssegaard, C. B., Hennestad, B. W., Wang, M. V., Martinsen, J.,…Johnsen, T. (2018). *Inkluderende fellesskap for barn og unge. Ekspertgruppen for barn og unge med behov for særskilt tilrettelegging*. Oslo: Fagbokforlaget, Vigmostad & Bjørke AS.

Nordahl, T., & Sunnevåg, A-K. (2008). *Spesialundervisningen i grunnskolen - stor avstand mellom idealer og realiteter*. Rapport 2 - 2008. Elverum: Høgskolen i Hedmark.

Nordahl, T, Sørlie, M-A, Tveit, A og Manger, T. (2003a): *Alvorlige atferdsvansker – effektiv forebygging og mestring i skolen. Veileder for lærere*. Oslo: Læringssenteret.

Nordahl, T, Sørlie, M-A, Tveit, A og Manger, T. (2003b): *Alvorlige atferdsvansker – effektiv forebygging og mestring i skolen. Veileder for skoleeier og skolens ledelse*. Oslo: Læringssenteret.

Nordahl, Sørlie, Manger, & Tveit (2005). *Atferdsproblemer blant barn og unge. Teoretiske og praktiske tilnærminger*. Bergen: Fagbokforlaget.

Nyskole (2008). *En kritikk av atferdsprogrammet PALS i skolen.* 03.01.08. http://www.forumnyskole.org/en-kritikk-av-adferdsprogrammet-pals-i-skolen/

O

Oakes, W. P., Lane, K. L., Jenkins, A., & Broker, B. B. (2013). Three-tiered models of prevention: Teacher efficacy and burnout. *Education & Treatment of Children, 36*, 95-126.

OECD (2007). *Knowledge management. Evidence in education. Linking research and policy*. Paris: OECD publishing/Center for Educational Research and Innovation (CERI).

OECD (2014). *Talis 2013 Results: An international perspective on teaching and learning*. TALIS,

OECD Publishing. http://dx.doi.org/10.1787/9789264196261-en

OECD (2014). *Education at a Glance 2014: OECD Indicators.* OECD Publishing. http://dx.doi.org/10.1787/eag-2014-en

OECD (2018). The gender gap in education. Outcomes in Norway. OECD Education Working Paper No. 183. OECD, Directorate for Education and Skills. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=EDU/WKP(2018)19&docLanguage=En

Official Directive Q-16 (2007). *Fokus på forebyggende innsatser for barn og unge*. Oslo: Barne- og likestillingsdepartementet.

Official Report no. 19 (NOU 19:2009). *Rett til læring*. Oslo: Kunnskapsdepartementet.

Official Report no.7 (NOU 7:2010). *Mangfold og mestring. Flerspråklige barn, unge og voksne i opplæringssystemet*. Oslo: Kunnskapsdepartementet.

Official Report no. 2 (NOU 2:2015). *Å høre til. Virkemidler for et trygt psykososialt skolemiljø.* Oslo: Kunnskapsdepartementet.

Official Report no. 8 (NOU (8:2015). *Fremtidens skole. Fornyelse av fag og kompetanser*. Oslo: Kunnskapsdepartementet.

Official Report no. 3 (NOU 3: 2019). *Nye sjanser – bedre læring. Kjønnsforskjeller i skoleprestasjoner og utdanningsløp*. Oslo: Departementenes sikkerhets- og serviceorganisasjon. Teknisk redaksjon.

Office of Juvenile Justice and Delinquency Prevention (OJJDP). (1995). *Guide for implementing the comprehensive strategy for serious, violent, and chronic juvenile offenders*. Washington, DC: U.S. Department of Justice, Office of Juvenile Justice and Delinquency Prevention.

Ogden, T. (1990). *Kvalitetsbevissthet i skolen.* Oslo: Universitetsforlaget.

Ogden, T. (1992): *Atferdspedagogikk i teori og praksis. Om arbeid med problematferd i skolen*. Universitetsforlaget.

Ogden, T. (1995). *Kompetanse i kontekst. En studie av risiko og kompetanse hos 10 og 13 åringer. Prosjekt Oppvekstnettverk*. Rapportserie fra Barnevernets Utviklingssenter, nr. 3. Oslo: Barnevernets Utviklingssenter.

Ogden, T. (1998). *Elevatferd og læringsmiljø. Lærernes erfaringer med og syn på elevatferd og læringsmiljø.* Oslo: Kirke-, utdannings-, og forskningsdepartementet.

Ogden, T. (2000). Elevforskjeller som utfordring for enhetsskolen. *Bedre Skole, 1*, 6-15.

Ogden, T. (2003). The validity of teacher ratings of adolescents' social skills. *Scandinavian Journal of Educational Research, 47, 63-76.*
Ogden, T. (2008). Motviljen mot evidens i utdanningssystemet. *Bedre Skole, 4,* 75-79.

Ogden, T. (2009). *Sosial kompetanse og problematferd i skolen. 2. utg.* Oslo: Gyldendal Akademisk.

Ogden, T. (2010). *Familiebasert behandling aalvorlige atferdsproblemer blant barn og unge. Evaluering og implementering av evidensbaserte behandlingsprogrammer i Norge*. Avhandling dr. Philos. Berge: Universitetet i Bergen

Ogden, T. (2012). *Evidensbasert praksis i arbeidet med barn og unge*. *Oslo: Gyldendal Akademisk.*

Ogden, T. (2014). Special needs education in Norway - the past, present, and future of the field. In B. G. Cook, M. Tankersley & T. J. Landrum (eds.): *The past, the present and the future of special education: Advances in learning and behavioral disabilities*, vol. 27 (pp. 213-238). New York: Emerald Group Publishing ltd.

Ogden, T. (2015). *Sosial kompetanse og problematferd blant barn og unge*. 3. utg. Oslo. Gyldendal Akademisk

Ogden , T. , & Hagen , K. A. (2008). Treatment effectiveness of Parent Management Training in Norway: A randomized controlled trial of children with conduct problems. *Journal of Consulting and Clinical Psychology, 76*, 607 – 621.

Ogden, T., & Hagen, K. A. (2014). *Adolescent mental health. Prevention and intervention*. London: Routledge.

Ogden, T., & Halliday-Boykins, C. (2004). Multisystemic treatment of antisocial adolescents in Norway: Replication of clinical outcomes outside of the US. *Child and Adolescent Mental Health, 9*, 77-83.

Ogden, T., Sørlie, M-A., Arnesen, A., & Meek-Hansen, W. (2012). The N-PALS School-wide positive behaviour support model in Norwegian primary schools - implementation and evaluation. In Visser, J., Daniels, H., & Cole, T. (Eds.). Transforming troubled lives: Strategies and interventions for children with social, emotional and behavioural difficulties. Vol. 2. International perspectives on inclusive education (pp. 39-55). *Oxford: Emerald Group Publishing Limited.*

Ogden, T., Sørlie, M-A., & Hagen, K. A. (2007). Building strength through enhancing social competence in immigrant students in primary school. A pilot study. *Journal of Emotional and Behavioural Difficulties, 12*, 105-117.

Ogders, C. L., Moffitt, T. E., Broadbent, J. M. (2008). Female and male antisocial trajectories: From childhood origins to adult outcomes. *Development and Psychopathology, 20*, 673-716.

Olweus, D. (1991). Bully/victim problems among schoolchildren: Basic facts and effects of a school based intervention program. In D. J. Pepler & K. H. Rubin (Eds.), *The development and treatment of childhood aggression.* Hillsdale, NJ: Erlbaum.

Olweus, D., & Limber, S. P. (2010). Bullying in school: Evaluation and dissemination of the Olweus Bullying Prevention Program. *American Journal of Orthopsychiatry*, *80*, 124-134.

Oliver, R. M., Wehby, J. H., & Reschly, D. J. (2011). Teacher classroom management practices: Effects on disruptive or aggressive student behavior. *Campbell Systematic Reviews*, *4*, 1-55.

OSEP Technical Assistance Center on Positive Behavioral Interventions and Supports (October 2015). *Positive Behavioral Interventions and Supports (PBIS) Implementation Blueprint: Part 1 – Foundations and Supporting Information.* Eugene, OR: University of Oregon. www.pbis.org.

OSEP (Center on Positive Behavioral Interventions and Supports) (2007). *Is school-wide positive behavior support an evidence-based practice? A research summary*. www.pbis.org.

Oyen, K. A., & Wollersheim-Shevey, S. (2018). An examination of critical features of positive frameworks: Impact in rural environments for school-based practitioners. *Contemporary School Psychology*. https://doi.org/10.1007/s-40688-018-0198-6

P

Parker, R. I., Brossart, D. F., Vannest, K. J., Long, J. R., De-Alba, R. G., Baugh, F. G., et al. (2005). Effect size in single-case research. How large is large? *School Psychology Review, 34,* 116-132.

Pas, E. T., & Bradshaw, C. P. (2012). Examining the association between implementation and outcomes: State-wide scale-up of school-wide positive behavior intervention and supports. *Journal of Behavioral Health Services & Research, 39*, 417–433.

Patterson, G. R. (1982). *A social learning approach*. *Coercive family process*. *Vol. 3*. Eugene, OR: Castalia Publishing Company.

Patterson, G. R., Reid, J. B., & Dishion, T. J. (1992). *Antisocial boys*. Eugene, OR: Castalia.

Paulus, P. (2009). Mental health - backbone of the soul. *Health Education, 109*, 289–298.

Pauwels, L. J. R., & Svensson, R. (2015). Schools and child antisocial behavior. In search for mediator effects of school-level disadvantage. *Sage Open, April-June*, 1-13.

Payne, A. S. (2008). A multilevel analysis of the relationships among communal school organization, student bonding, and delinquency. *Journal of Research in Crime and Delinquency, 45,* 429-455.

Payne, C. M. (2008). *So much reform, so little change: The Persistence of failure in urban schools*. Cambridge, MA: Harvard Education Press.

Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology, 48*, 85–112.

Pianta, R. C., Belsky, J., Vandergrift, N., Houts, R., & Morrison, F. J. (2008). Classroom effects on childrens achievement trajectories in elementary school. *American Educational Research Journal, 45*, 365-397.

Pittman, R. B., & Haughwout, P. (1987). Influence of high school size on dropout rate. *Educational Evaluation and Policy Analysis, 9*, 337-343.

Pogow, S. (2019). How effect size (practical significance) misleads clinical practical practice: The case for switching to practical benefit to assess applied research findings. *The American Statistican, 73:Sup1*, 223-234.

Prentice, D. A., & Miller, D. T. (1992). When small effects are impressive. *Psychological Bulletin, 112*, 160-164.
Publications APA, on Journal, C. B. W. G., & Standards, A. R. (2008). Reporting standards for research in psychology: why do we need them? What might they be? *The American Psychologist*, *63*, 839-851.

Putallaz, M., & Bierman, K. A. (Eds.) (2004). *Aggression, antisocial behavior, and violence among girls: A developmental perspective.* New York: Guilford Press.

Putnan, R. F., & Kostner, T. (2015). A reply to the commentaries on "School-wide PBIS: An Example of Applied Behavior Analysis Implemented at a Scale of Social Importance" by Horner and Sugai (2015): PBIS is function over form: The clear behavioral roots and opportunities the PBIS framework presents to the field of behavior analysis moving forward. ***Behavior Analysis Pract****ice*. Doi 10.1007/s40617-015-0092-x

R

Ragsdale, K. (1996). *Selection bias in nonrandomized experiments: An investigation using meta-analysis* (Unpublished doctoral dissertation). Memphis, TN: The University of Memphis.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models*. Newbury Park, CA: Sage.

Raver, C. C., Jones, S. M., Li-Grining, C .P., Metzger, M., Champion, K. M., & Sardin, L. (2008). Improving preschool classroom processes: Preliminary findings from a randomized trial implemented in Head Start settings. *Early Childhood Research Quarterly, 23*, 10–26.

Reaves, S., McMahon, S. D., Duffy, S. N., & Ruiz, L. (2018). A test of time: a meta-analytic review of the relation between school climate and problem behavior. *Aggression and Violent Behavior, 39,* 100-108.

Reef, J., Diamantopoulou, S., van Meurs, I., Verhulst, F. C., & van der Ende, J. (2011). Developmental trajectories of child to adolescent externalizing behavior and adult DSM-IV disorder: results of a 24-year longitudinal study. *Social Psychiatry and Epidemiology 46,* 1233-1241.

Reid, J. B. (1993). Prevention of conduct disorder before and after school entry: Relating interventions to developmental findings. *Development and Psychopathology*, 5, 243-262.

Reid, J. B., Patterson, G. R. & Snyder, J. (eds.) (2002). *Antisocial behavior in children and adolescents. A developmental analysis and model for intervention*. Washington DC, American Psychological Association.

Reinke, W. M., Herman, K. C., & Stormont, M. (2013). Classroom-level positive behavior supports in schools implementing SW-PBIS identifying areas of enhancement. *Journal of Positive Behavior Interventions, 15,* 39-50.

Report to the Storting no. 28 (1998-1999). *Mot rikare mål. Om einskapsskolen, det likeverdige opplæringstilbodet og ein nasjonal strategi for vurdering og kvalitetsutvikling i grunnskolen og den vidaregåande opplæringa*. Oslo: Kunnskapsdepartementet.

Report to the Storting no. 17 (1999-2000). *Handlingsplan mot barne- og ungdomskriminalitet*. Oslo: Ministry of Justice and Police.

Report to the Storting no. 30 (2003-2004). *Kultur for læring*. Oslo: Kunnskapsdepartementet.

Report to the Storting no. 16 (2006-2007). *…og ingen sto igjen – tidlig innsats for livslang læring*. Oslo: Kunnskapsdepartementet.

Report to the Storting no.31 (2007-2008). *Kvalitet i skolen*. Oslo: : Kunnskapsdepartementet.

Report to the Storting no.18 (2010-2011*). Læring og fellsskap. Tidlig innsats og gode læringsmiljøer*. Oslo: Kunnskapsdepartementet.

Report to the Storting no. 21 (2016-2017). *Lærelyst – tidlig innsats og kvalitet i skolen.* Oslo: Kunnskapsdepartementet.

Report to the Storting no. 6. (2019-2020). *Tett på – tidlig innsats og inkluderende fellesskap i barnehage, skole og SFO*. Oslo: Kunnskapsdepartementet.

Rescola, L., Achenbach, T. M., Ivanova, M. Y., Dumenci, L., Almqvist, F., Bilenberg, N., et al. (2007). Epidemiological comparisons of problems and positive qualities reported by adolescents in 24 countries. *Journal of Consulting and Clinical Psychology, 75Re,* 351–358.

Rescorla, L. A., Ginzburg, S., Achenbach, T. M., Ivanova, M. Y., Almqvist, F., Begovac, I., ….Verhulst, F. (2013). Cross-informant agreement between parent-reported and adolescent self-reported problems in 25 societies. *Journal of Clinical Child & Adolescent Psychology, 42,* 262-273.

Rimfeld, K., Malanchini, M., Hannigan; L. J., Dale, P. S., Allen, R., Hart, S. A., & Plomin, R. (2019). Teacher assessments during compulsory education are as reliable, stable and heritable as standardized test scores. *Journal of Child Psychology and Psychiatry, 60,* 1278-1288.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review, 15,* 351–357.

Romano, E., Weegar, K., Babchishin, L., & Saini, M. (2018). Cross-informant agreement on mental health outcomes in children with maltreatment histories: A systematic review. *Psychology of Violence, 8,* 19–30.

Rosnow, R., & Rosenthal, R. (2003). Effect sizes for experimenting psychologists. *Canadian Journal of Experimental Psychology, 57,* 221-237.

Ross, S. W., & Horner, R. H. (2007). Teacher outcomes of School-Wide Positive Behavior Support. *Teaching Exceptional Children Plus, 3,* Article 6. http://escholarship.bc.edu/education/tecplus/vol3/iss6/art6

Ross, S. W., & Horner, R. H. (2009). Bully prevention in Positive Behavior Support. *Journal of Applied Behavior Analysis,* 42, 747-759.

Ross, S. W., Romer, N, & Horner, R. H. (2012). Teacher well-being and the implementation of school-wide positive behavior support: A case study. *Journal of Positive Behavior Support, 14,* 118-128.

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation. A systematic approach (7th ed.).* Thousand Oaks, CA: Sage.

Rucinski, C. L., Brown, J. L., & Dower, J. T. (2018). Teacher-child relationships, classroom climate, and children's social-emotional and academic development. *Journal of Educational Psychology, 110,* 992-1004.

Rutter, M. (1993). Resilience. Some conceptual considerations. *Journal of Adolescent Health Care, 14,* 626-631.

Ruzek, E. A., Domina, T., Conley, A. M., Duncan, G. J., & Karabenick, S. A. (2015). Using value-added models to measure teacher effects on students' motivation and achievement. *Journal of Early Adolescence, 35,* 852-882.

S

Safran, S. P., & Oswald, K. (2003). Positive behavior supports. Can schools reshape disciplinary practices? *Exceptional Children, 69,* 361-373.

Sailor, W., Zuna, N., Choi, J-H., Thomas, J., McCart, A., & Roger, B. (2006). Anchoring Schoolwide Positive Behavior Support in structural school reform. *Research & Practice for Persons with Severe Disabilities, 31,* 18-30.

Sameroff, A. J., & Mackenzie, M. J. (2003). Research strategies for capturing transactional models of development: The limits of the possible. *Development and Psychopathology, 15,* 613-640.

Sammons, P., Thomas, S., & Mortimore, P. (1997). *Forging links: Effective departments and effective schools*. London: Paul Chapman.

Sampson, R. J., & Laub, J. H. (2005). A life-course view of the development of crime. *The Annals of the American Academy of Political and Social Science, 602,* 12-45.

Sanders, E. A. (2011). *Multilevel analysis methods for partially nested cluster randomized trials*. Dissertation, Doctor of Philosophy, College of Education, University of Washington.

Santagata, R., Kersting, N., Givvin, K. B., & Stigler, J. W. (2011). Problem implementation as a lever for change: An experimental study of the effects of a professional development program on students' mathematics learning. *Journal of Research on Educational Effectiveness, 4*, 1–24.

Saretsky, G. (1972). The OEO P.C. experiment and the John Henry Effect. The Phi Delta Kappan. 53, 579–581.

Schafer, J. L., & Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychological Methods, 7*, 147-177.

Scheckner, S., Rollin, S., Kaiser-Ulrey, C., & Wagner, R. (2002). School violence in children and adolescents: a meta-analysis of the effectiveness of current interventions. J*ournal of School Violence, 1,* 5–33.

Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin College Div.

Shin, H., & Ryan, A. M. (2017). Friend influence on early adolescent disruptive behavior in the classroom: Teacher emotional support matters. *Developmental Psychology, 53*, 114–125.

Schools for Health in Europe (2010). http://ws10.e-vision.nl/she_network/index.cfm
Schloss, P. J., & Smith, M. A. (1998). *Applied behavior analysis in the classroom*. Pearson College Division.

Scott, T. M. (2001). A schoolwide example of positive behavior support. *Journal of Positive Behavior Interventions, 3*, 88-95.

Scott, T. M., & Eber, L. (2003). Functional assessment and wraparound as systemic school processes: Primary, secondary, and tertiary systems examples. *Journal of Positive Behavior Interventions, 5*, 131–14.

Shadish, W. R. (2011). Randomized controlled studies and alternative designs in outcome studies: Challenges and opportunities. *Research on Social Work Practice, 21*, 636-643.

Shadish, W. R., Clark, M. H., & Steiner, P. M. (2008). Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments. *Journal of the American Statistical Association, 103,* 1334-1344.

Shadish, W. R., & Cook, T. D. (2009). The renaissance of field experimentation in evaluating interventions. *Annual review of psychology, 60*, 607-629.

Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin College Div.

Shadish, W. R., Galindo, R., Wong, V. C., Steiner, P. M., & Cook, T. D. (2011). A randomized experiment comparing random and cutoff-based assignment. *Psychological methods, 16,* 179-191.

Shadish, W. R., Navarro, A. M., Matt, G. E., & Phillips, G. (2000). The effects of psychological therapies under clinically representative conditions: a meta-analysis. *Psychological bulletin, 126*, 512-529.

Shadish, W. R., & Ragsdale, K. (1996). Random versus nonrandom assignment in controlled experiments: Do you get the same answer? *Journal of consulting and clinical psychology, 64*, 1290-1305.

Shaw, S. M., Walls, S. M., Dacy, B. S., Levin, J. R., & Robinson, D. H. (2010). A follow-up note on prescriptive statements in nonintervention research studies. *Journal of Educational Psychology, 102,* 982-988.

Shirley, E. M., & Cornell, D. G. (2012). The contribution of student perceptions of school climate to understanding the disproportionate punishment of African American students in a middle school. *School Psychology International, 33*, 115-134.

Shucksmith, J., Summerbell, C., Jones, S., & Whittaker, V. (2007) *Mental wellbeing of children in primary education (targeted/indicated activities).* London: National Institute of Clinical Excellence.

Silberg, J., Moore, A. A., & Rutter, M. (2015). Age of onset and the subclassification of conduct/ dissocial disorder. *Journal of Child Psychology and Psychiatry, 56*, 826–833.

Simonsen, B., Eber, L., Black, A. C., Sugai, G., Lewandowski, H., Sims, B., & Meyers, D. (2012). Illinois statewide positive behavioral interventions and supports evolution and impact on student outcomes across years. *Journal of Positive Behavior Interventions, 14*, 5-16.

Singer, J. D., & Willet, J. B. (2003). A framework for investigating change over time. In *Applied longitudinal data analysis: Modeling change and event occurrence (pp.* 3-15). New York: Oxford University Press.

Sink, C., & Stroh, H. (2006). Practical significance: The use of effect sizes in school counseling research. *Professional School Counseling, 9*, 401-411.

Sklad, M., Diekstra, R., De Ritter, M., Ben, J., & Gravestejn, C. (2012). Effectiveness of school-based universal social, emotional, and behavioral programs: Do they enhance students' development in the area of skill, behavior, and adjustment? *Psychology in the Schools, 49,* 892-909.

Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis.* New York: D. Appleton-Century Company.

Skogen, J. C., & Torvik, F. A. (2013). *Atferdsforstyrrelser blant barn og unge i Norge: Beregnet forekomst og bruk av hjelpetiltak.* Rapport 2013:4. Oslo: Folkehelseinstituttet.

Slavin, R. E. (2003). *Educational psychology theory and practice* (7th ed). New York: Pearson Education, Inc.

Solomon, B. G., Kein, S. A., Hintze, J. M., Cressy, J. M., & Peller, S. L. (2012). A meta-analysis of School-Wide Positive Behavior Support. An exploratory study using single-case synthesis. *Psychology in the Schools, 49,* 105-121.

Solomon, D., Watson, M., Battistich, V., Schaps, E., & Delucchi, K. (1996). Creating classrooms that students experience as communities. *American Journal of Community Psychology, 24,* 716-748.

Soukamneuth, S. (2004). A climate for learning. *Principal Leadership; High School Edition, 4*, 14-19.

Sprague, J. R., Walker, H., Golly, A., White, K., Myers, D. R., & Shannon, T. (2002). Translating research into effective practice: The effects of a universal staff and student intervention on key indicators of school safety and discipline. *Education and Treatment of Children, 24*, 495-511.

Sroufe, L. A., & Rutter, M. (1984). The domain of developmental psychopathology. *Child Development*, 17-29.

Statistics Norway (2009). *Education statistics. Elementary and lower secondary schools.* http://www.ssb.no/english/subjects/04/02/20/utgrs_en/tab- 2009-04-29-05-en.html

Statistics Norway (SSB, 2017). *Pupils in primary and lower secondary school.* Table 1. Primary and lower secondary schools. Pupils in primary and lower secondary schools. https://www.ssb.no/utdanning/statistikker/utgrs/aar

Stattin, H. & Magnusson, D. (1995). Onset of official delinquency: its co-occurrence in time with educational, behavioural and interpersonal problems. *British Journal of Criminology*, 35, 417-449.

Steffgen, G., Recchia, S., & Viechtbauer, W. (2013). The link between school climate and violence in school: A meta-analytic review. *Aggression and Violent Behavior, 18*, 300-309.

Steiner, P. M., Cook, T. D., Shadish, W. R., & Clark, M. H. (2010). The importance of covariate selection in controlling for selection bias in observational studies. *Psychological methods, 15*, 250-267.
Stewart, E. A. (2003). School social bonds, school climate, and school misbehavior. A multi-level analysis. *Justice Quarterly, 20*, 575-601.

Stewart, R. M., Benner, G. J., Martella, R. C., & Marchand-Martella, N. E. (2007). Three-tier models of reading and behavior. A research review. *Journal of Positive Behavior Interventions, 9*, 239-253.

Sugai, G., & Horner, R. H. (1994). Including students with severe behavior problems in general education settings: Assumptions, challenges, and solutions. In J. Marr, G. Sugai, & G. Tindal (Eds.). *The Oregon conference monograph (Vol. 6)*.

Sugai, G., & Horner, R. (2002). The evolution of discipline practices: School-Wide Positive Behavior Supports. *Child & Family Behavior Therapy, 24*, 23-55.

Sugai, G., Horner, R. H., Algozzine, R., Barrett, S., & Lewis, T., Anderson, C….Simonsen, B. (2010). *School-wide positive behavior support: Implementers' blueprint and self-assessment.* Eugene, OR: University of Oregon.

Sugai, G., Horner, R. H., Algozzine, R., Barrett, S., Lewis, T., Anderson, C. ...Simonsen, B. (2015). *School-wide positive behavior support: Implementers' blueprint and self-assessment*. Eugene, OR: University of Oregon. https://www.pbis.org/blueprint/implementation-blueprint

Sugai, G., Horner, R. H., Dunlap, G., Hieneman, M., Lewis, T. J., Nelson, C. M., …..& Todd, A. W. (2000:2009). *Effective behavior support (EBS Survey). Assessing and planning behavior support in schools.* Eugene, OR: University of Oregon.

Sugai, G., Horner, R. H., & Todd, A. W. (2000). *Effective behavior support (EBS) self-assessment survey.* Eugene, OR: University of Oregon.

Sugai, G., Horner, R. H., & Todd, A. W. (2009). *Effective behavior support (EBS) self-assessment survey (Version 3).* Eugene, OR: University of Oregon, Educational and Community Supports. https://www.pbisapps.org/Applications/Pages/PBIS-Assessment-Surveys.aspx#sas

Sugai, G., Lewis-Palmer, T., & Hagan-Burke, S. (2000). Overview of the functional behavioral assessment process. *Exceptionality, 8*, 149–160.

Sugai, S., Sprague, J. R., Horner, R. H., & Walker, H. M. (2000). Preventing school violence. The use of office discipline referrals to assess and monitor school-wide discipline interventions. *Journal of Emotional and Behavioral Disorders, 8*, 94-101.

Sundell, K., & Stensson, E. (2010). *Effektutvärderingar i doktorsavhandlingar*. Stockholm: Socialstyrelsen.

Sørlie, M.-A. (2000). *Alvorlige atferdsproblemer og lovende tiltak i skolen*. Oslo: Praxis Forlag.

Sørlie, M-A., Hagen, K. A., & Nordahl, K. B. (2020). Development of social skills during middle childhood: Growth trajectories and school-related predictors. *International Journal of School & Educational Psychology*. Doi: 10.1080/21683603.2020.1744492

Sørlie, M-A., Hagen, K. A., & Ogden, T. (2008). Social competence and antisocial behavior: Continuity and distinctiveness across early adolescence. *Journal of Research on Adolescence, 18*, 121–144.

Sørlie, M.-A., & Nordahl, T. (1998). *Problematferd i skolen. Hovedfunn, forklaringer og pedagogiske implikasjoner. Hovedrapport fra forskningsprosjektet «Skole- og samspillsvansker»*. Rapport 12a/98. Oslo: Norsk institutt for forskning om oppvekst, velferd og aldring.

Sørlie, M-A., & Ogden, T. (2007). Immediate impacts of N-PALS: A school-wide multilevel programme targeting behaviour problems in elementary school. *Scandinavian Journal of Educational Research, 51*, 471–492.

Sørlie, M-A. & Ogden, T. (2014). Mindre problematferd i grunnskolen? Lærervurderinger i et 10-års perspektiv. *Norsk Pedagogisk Tidsskrift, 98*, 190-202.

Sørlie, M. A., & Ogden, T. (2014). Reducing threats to validity by design in a nonrandomized experiment of a school-wide prevention model. *International Journal of School & Educational Psychology, 2*, 235-246.

Sørlie, M-A., & Ogden, T. (2015). School-Wide Positive Behavior Support–Norway: Impacts on problem behavior and classroom climate. *International Journal of School & Educational Psychology, 3*, 202-217.

Sørlie, M-A., Ogden, T., Arnesen, A., Olseth, A. R. & Meek-Hansen, W. (2014). Skoleledere om PALS. En brukerundersøkelse om erfaringer med og vurderinger av en skoleomfattende tiltaksmodell for positiv læringsstøtte. *Spesialpedagogikk, 1*, 17-29.

Sørlie, M. A., Ogden, T., & Olseth, A. R. (2015). Preventing problem behavior in school through school-wide staff empowerment: Intervention outcomes. *World Journal of Educational Research*, *2*, 117-139.

Sørlie, M-A., Ogden, T., & Olseth, A. R. (2016). Examining teacher outcomes of the School-Wide Positive Behavior Support model in Norway: perceived efficacy and behavior management. *Sage Open, 00*, 1-13. Doi: 10.1177/2158244016651914

Sørlie, M-A., Ogden, T., Solholm, R., & Olseth, A. R. (2010). Implementeringskvalitet – om å få tiltak til å virke: En oversikt. *Tidsskrift for norsk psykologforening, 47*, 315–321.

Sørlie, M-A., & Torsheim, T. (2011). Multilevel analysis of the relationship between teacher

collective efficacy and problem behaviour in school, *School Effectiveness and School Improvement, 22*, 175-191.

T

Taylor-Greene, S., Brown, D., Nelson, L., Longton, J., Gassman, T., Cohen, J., Swartz, J., Horner, R. H., Sugai, G., & Hall, S. (1997). School-wide behavioral support: Starting the year off right. *Journal of Behavioral Education, 7*, 99-112.

Taylor-Greene & Kartub, D. (2000). Durable implementation of School-wide Behavior Support: The High Five Program. *Journal of Positive Behavior Interventions, 2*, 231-232.

Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro (2013). A review of school climate research. *Review of Educational Research, 83*, 357-385.

Thompson, B. (2002). "Statistical", "practical", and "clinical": How many kinds of significance do counselors need to consider? *Journal of Counseling and Development, 80*, 64-71.

Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology, 51*, 473-481.

Tobin, T. J., & Sugai, G. (2005). Preventing problem behaviours: Primary, secondary, and tertiary level prevention interventions for young children. *Journal of Early and Intensive Behavior Intervention, 3*, 125–144.

Todd, A., Campbell, A., Meyer, G., & Horner, R. (2008). Evaluation of a targeted group intervention in elementary students: The check-in/check-out program. *Journal of Positive Behavior Interventions. 10*, 46–55.

Todd, A. W., Horner, R. H., Sugai, G., & Sprague, J. R. (1999). Effective behavior support: Strengthening school-wide systems through a team-based approach. *Effective School Practices, 17*, 23-37.

Tolan, P., & Guerra, N. (1994). *What works in reducing adolescent violence? An empirical review of the field*. Boulder, CO: Center for the Study and Prevention of Violence. University of Colorado.

Toren, N. K., & Seginer, R. (2015). Classroom climate, parental educational involvement, and student school functioning in early adolescence: A longitudinal study. *Social psychology of education, 18*, 811-827.

Turnbull, A., Edmonson, H., Griggs, P., Wickham, D., Sailor, W., Freeman, R. et al. (2002). A blueprint for Schoolwide Positive Behavior Support. Implementation of three components. *Exceptional Children, 68*, 377-402.

U

Ukoumunne, O. C., Guilford, M. C., Chinn, S., Sterne, J. A. C., & Burney, P. G. J. (1999). *Methods for evaluating area-wide and organization-based interventions in health and health-care: A systematic review*. Southapton, UK: University of Southapton, Health Technology Assessment, 3(5).

Umbreit, J., & Blair, K-S. (1996). The effect of preference, choice, and attention on problem behavior at school. *Education and Training in Mental Retardation and Developmental Disabilities, 31*, 151-161.

UNICEF (1989). *Convention on the rights of children*. https://www.unhcr.org/uk/4d9474b49.pdf

UNESCO (1994). *The Salamanca Statement and Framework for Action on Special Needs Education.* Paris: UNESCO.

Utdanningsdirektoratet (2012a). *Skoleledere – implementering*. http://www.udir.no/Laringsmiljo/Klasseledelse/Sider/Modell-for-praktisk-utviklingsarbeid/Skoleleder---Implementering/?idkeep=True

Utdanningsdirektoratet (2012b). *Støtte til læringsmiljøer og antimobbeprogram*. http://www.udir.no/Laringsmiljo/Arbeid-mot-mobbing/Stotte-til-laringsmiljo--og-antimobbeprogrammer/

Utdanningsforbundet (2008). *Evidens og evidensdebattens betydning for utdanningssystemet*. Temanotat 6/2008. Oslo: Utdanningsforbundet.

V

van der Scheer, E. A., Bijlsma, H. J. E., & Glas, C. A. W. (2019). Validity and reliability of student perceptions of teaching quality in primary education. *School Effectiveness and School Improvement, 30*, 30-50.

Vasallo, S., Edvards, B., & Forest, W. (2016). Childhood behavior problems and fighting in early adulthood: What factors are protective? *Journal of Criminal Justice, 45*, 85-93.

Vygotskij, L. S. (Ed.). (1978). *Mind in society. Development of higher psychological processes.* Cambridge, Mass.: Harvard University Press.

W

Wagner, W., Göllner, R., Werth, S., Schmitz, B., & Voss, T. (2016). Student and teacher ratings of instructional quality: Consistency of ratings over time, agreement, and predictive power. *Journal of Educational Psychology, 108*, 705-721.

Walker, H., Colvin, G., & Ramsey, E. (1995). *Antisocial behavior in public school: Strategies and best practices*. Pacific Grove, CA: Brookes/Cole.

Walker, B., Cheney, D., Stage, S., Blum, C., & Horner, R. H. (2005). Schoolwide screening and positive behavior supports: Identifying and supporting students at risk for school failure. *Journal of Positive Behavior Interventions, 7*, 194–204.

Walker, H. M., Forness, S. R., Kauffman, J. M., Epstein, M. H., Gresham, F. M., Nelson, C. M., & Srrain, P. S. (1998). Macro-social validation. Referencing outcomes in behavioral disorders to social issues and problems. *Behavioral Disorders, 24*, 17-18.

Walker, H. M., Horner, R. H., Sugai, G., Bullis, M., Sprague, J. R., Brticker, D., & Kaufman, M. J. (1996). Integrated approaches to preventing antisocial behavior patterns among school-age children and youths. *Journal of Emotional and Behavioral Disorders, 4*, 194-209.

Walker, H. M., Zeller, R. W., Close, D. W., Webber, J., & Gresham, F. (1999). The present unwrapped. Change and challenge in the field of behavioral disorders. *Behavioral Disorders, 24*, 293-304.

Walliman, N. (Ed.) (2006). Social research methods. London: Sage.

Wampold, B. E., & Serlin, R. C. (2000). The consequence of ignoring a nested factor on measures of effect size in analysis of variance. *Psychological Methods, 5*, 425-433.

Wang, M. (2009). School climate support for behavioral and psychological adjustment. Testing the mediating effect of social competence. *School Psychology Quarterly, 24*, 240-251.

Wang, M-T., & Degol, J. L. (2016). School climate. A review of the construct, measurement, and impact on student outcomes. *Educational Psychology Review, 28*, 315–352.

Wang, M-T., Selman, R. L., Dishion, T. J., & Stormshak, E. A. (2010).A Tobit regression analysis of the covariation between middle school students perceived school climate and behavioral problems. *Journal of Research on Adolescence, 20*, 274-286.

Waasdrop, T. E., Bradshaw, C. P., & Leaf, P. J. (2012). The impact of schoolwide positive behavioral interventions and supports on bullying and peer rejection: A randomized controlled effectiveness trial. *Archives of Pediatrics & Adolescent Medicine, 166*, 149-156.

Weare, K. (2015). *What works in promoting social and emotional well-being and responding to mental health problems in schools?* London: National Children's Bureau.

Weare, K., & Nind, M. (2011). Mental health promotion and problem prevention in schools. What does evidence say? *Health Promotion International, 26, S*, i29-i69.

Weinstein, R. S. (1985). Student mediation of classroom expectancy effects. In J. B. Dusek (Ed.), *Teacher expectancies* (pp. 329-350). Hillsdale, NJ: Erlbaum.

Welsh, W. N., Greene, J. R., & Jenkins, P. H. (1999). School disorder. The influence of individual, institutional, and community factors. *Criminology, 37*, 73-116.

Wells, J., Barlow, J., & Stewart-Brown, S. (2003). A systematic review of universal approaches to mental health promotion in schools. *Health Education, 103*, 197–220.

Wendelborg, C., Paulsen, V., Røe, M., Valenta, M, & Skaalvik, E. (2012). *Elevundersøkelsen 2012. Analyse av Elevundersøkelsen 2012*. Rapport 2012. Mangfold og inkludering. Trondheim: NTNU Samfunnsforskning as.

Wendelborg, C., Caspersen, J., & Kongsvik, T. (2015). *Mot et større mangfold? Systemrettet arbeid og tilpasset opplæring i kommuner med lavt omfang av spesialundervisning*. Rapport 2015.Mangfold og inkludering. Trondheim: NTNU Samfunnsforskning as.

Weng, M-T., & Degol, J. L. (2016). School climate. A review of the construct, measurement, and impact on student outcomes. *Educational Psychological Review, 28*, 315-352.

Weisburd, D. (2010). Justifying the use of non-experimental methods and disqualifying the use of randomized controlled trials: challenging folklore in evaluation research in crime and justice. *Journal of Experimental Criminology, 6*, 209-227.

Welsh, W. N. (2003). Individual and institutional predictors of school disorder. *Violence and Juvenile Justice, 4*, 346-368.

Westen, D., Novotny, C. M., & Thompson-Brenner, H. (2004). The empirical status of empirically supported psychotherapies: assumptions, findings, and reporting in controlled clinical trials. *Psychological Bulletin, 130*, 631-663.

White paper no. 28 (1998-1999). *Mot rikare mål. Om einskapsskolen, det likeverdige opplæringstilbodet og ein nasjonal strategi for vurdering og kvalitetsutvikling i grunnskolen og den vidaregåande opplæringa*. Oslo: Kirke-, utdannings- og forskningsdepartementet.

White paper no. 17 (1999-2000). *Handlingsplan. Sammen mot barne- og ungdomskriminalitet*. Oslo: Justis- og politidepartementet.

White Paper no. 30 (2003-2004). *Kultur for læring*. Oslo: Utdannings- og forskningsdepartementet.

White Paper no.16 (2006-2007). *…og ingen sto igjen – tidlig innsats for livslang læring*. Oslo: Kunnskapsdepartementet.

White Paper no.31 (2007-2008). *Kvalitet i skolen*. Oslo: Kunnskapsdepartementet.

White Paper no.18 (2010-2011*). Læring og fellsskap. Tidlig innsats og gode læringsmiljøer*. Oslo: Kunnskapsdepartementet.

Wilson, D. B., & Lipsey, M. W. (2001). The role of method in treatment effectiveness research: Evidence from meta-analysis. *Psychological Methods, 6*, 413-429.

Wilson, S. J., & Lipsey, M. W. (2007). School-based interventions for aggressive and disruptive behavior. Update of a meta-analysis. *American Journal of Preventive Medicine, 33*, 130-143.

Wilson, S. J., Lipsey, M., & Derzon, J. H. (2003). The effects of school-based intervention programs on aggressive behavior. A meta-analysis. *Journal of Consulting and Clinical Psychology, 71*, 136-149.

Wilcox, B. (2000). Applying positive behavior support and functional behavior assessment in schools. *Journal of Positive Behavioral Interventions, 2*, 131-143.

Wilson, A. W. (2015). A critique of sociocultural values in PBIS. *Behavior Analysis in Practice, 8*, 92-94.

Wineman, N. M., & Durand, E. (1992). Incentives and rewards for subjects in nursing research. *Western Journal of Nursing Research, 14*, 526-531.

Wissink, I. B., Dekovic´, M., Stams, G-J., Asscher, J. J., Rutten, E., & Zijlstra, B. J. H. (2014). Moral orientation and relationships in school and adolescent pro- and antisocial behaviors. A multilevel study. *The Journal of School Nursing, 30*, 216-225.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock-Johnson–Revised tests of achievement.* Chicago: Riverside.

Z

Zahn-Waxler, C., Schirtcliff, E. A., & Marceau, K. (2008). Disorders of childhood and adolescence: Gender and psychopathology. *Annual Review of Clinical Psychology, 4*, 275-303.

7. Paper 1-5

I

II

III

Article

SAGE Open
April-June 2016: 1–13

DOI: 10.1177/2158244016651914
sgo.sagepub.com
SAGE

Examining Teacher Outcomes of the School-Wide Positive Behavior Support Model in Norway: Perceived Efficacy and Behavior Management

Mari-Anne Sørlie[1], Terje Ogden[1], and Asgeir Røyrhus Olseth[1]

Abstract

Research on teacher outcomes of the School-Wide Positive Behavior Support (SWPBS) model has been scarce. The present study adds to the knowledge base by examining the effects of the Norwegian version of SWPBS (N-PALS) on school staffs' behavior management practices and on their individual and collective efficacy. Questionnaire data were collected from staff and students (Grades 4-7) at four measurement points across four successive school years in 28 intervention schools and 20 comparison schools. Using longitudinal multilevel analyses, indications of positive 3-year main effects of the N-PALS model were observed for staff-reported collective efficacy, self-efficacy, and positive behavior support practices. The intervention effects as measured by Cohen's *d* ranged from .14 to .91. The effects on student perceptions of teachers' behavior management strategies were, however, not consistent with the positive staff ratings. Results are discussed in relation to prior research, future research, and study limitations.

Keywords

crime prevention, educational research, collective behavior, SWPBS, school-wide intervention, effectiveness study, multilevel analysis

Important relations have been documented between teachers' behavior and efficacy perceptions and students' learning and conduct. Prior studies have indicated that low levels of student problem behavior, proactive classroom management, and a positive school climate support teachers' emotional wellbeing and their sense of efficacy, which in turn facilitate students' academic and social-emotional learning (e.g., Oliver, Wehby, & Reschly, 2011). Teachers' ability to organize classrooms and manage the behavior of their students is associated with high perceived efficacy and perseverance and appears to induce positive academic and psycho-social outcomes for students (Gibbs & Powell, 2011). Conversely, teachers struggling with classroom discipline often report high levels of stress and are frequently ineffective (Brouwers & Tomic, 2000).

A promising approach in the endeavor to prevent and manage student misconduct is the School-Wide Positive Behavior Support Model (SWPBS; Sprague & Walker, 2005), in which promoting staff efficacy, establishing more systematic and collective approaches, and more inclusive and supportive learning environments are key priorities. Since 2002, SWPBS has been implemented in more than 200 Norwegian elementary schools (7.5%) under the acronym N-PALS. SWPBS is a longitudinal and demanding model to implement, and to maintain the school staff's motivation and effort to implement the model with high fidelity across time it is essential to experience positive impacts not only for students but for staff as well. Accordingly, examining potential *teacher outcomes* is important in its own right, although research on teacher outcomes of SWPBS has been scarce (Ross, Romer, & Horner, 2012). The present study contributes to the knowledge base by examining the effects of N-PALS on school staffs' behavior management and on staffs' perceived individual and collective efficacy.

Teacher Self-Efficacy and Collective Efficacy

Based on social cognitive theory, it has been proposed that teachers' sense of individual and collective efficacy promotes positive teaching practices that in turn result in

[1]Atferdssenteret, Oslo, Norway

Corresponding Author:
Mari-Anne Sørlie, The Norwegian Center for Child Behavioral Development, Atferdssenteret, P.O. Box 7053 Majorstuen, 0306 Oslo, Norway.
Email: m.a.sorlie@atferdssenteret.no

enhanced student learning (Klassen & Tze, 2014). Within the school context, the construct of *self-efficacy* has been defined as the individual teacher's belief in his or her capability to affect student outcomes, particularly for unmotivated or difficult-to-teach students (Tschannen-Moran, Woolfolk Hoy, & Hoy, 2008). Teachers with a strong belief in their own efficacy are assumed to set more challenging goals, take more personal responsibility for student outcomes, and be more open to new teaching methods. Moreover, such teachers are more likely to be successful in implementing new programs, to be motivated to cope effectively with student misconduct, and to successfully manage challenges in the classroom (e.g., Tschannen-Moran et al., 2008). C*ollective efficacy* refers to school staffs' shared beliefs regarding their combined ability to organize and execute courses of action required to produce student success (Goddard, Hoy, & Woolfolk Hoy, 2004). Teachers' collective efficacy and teachers' self-efficacy are related but distinct constructs that vary across schools (e.g., Gibbs & Powell, 2011).

Relations Between Teacher Efficacy and Student Behavior

The majority of studies on efficacy in schools have focused on the relation between teacher self-efficacy and student achievement while there is little empirical evidence regarding how teachers' efficacy beliefs influence student behavior (Gibbs & Powell, 2011). A small-scale Israeli study indicated a significant positive relation between teacher self-efficacy and responses to student classroom behavior (Almog & Shechtman, 2007). Jordan and Stanovich (2003) found that teachers who expressed lower self-efficacy appeared less tolerant of problem behavior and unusual learning patterns and more likely to use punitive and reactive discipline strategies and seek exclusion of "problematic" students. As concerns the relationship between student problem behavior and teacher collective efficacy, Gibbs and Powell (2011) found that "efficacy for addressing external influences" as a component of the teachers' collective efficacy beliefs was significantly and inversely associated with the number of students excluded because of problem behavior. In a Norwegian study, Sørlie and Torsheim (2011) found a strong inverse but also reciprocal (bi-directional) relationship between collective efficacy and the level of problem behavior in schools, both concurrently and over time.

Effects of SWPBS on Teacher Efficacy

A few studies examining the effects of SWPBS on teacher self-efficacy have been published although no prior effectiveness study has examined the effects on collective efficacy in schools. Kelm and McIntosh (2012) examined aspects of self-efficacy in a controlled study and identified a large, significant effect of SWPBS on teachers' self-efficacy when controlling for school effects (d = .80). Ross and Horner (2007) found a significant effect on teachers' self-efficacy in SWPBS schools with high fidelity scores (n^2= .29). A recent study also reported significantly higher perceived self-efficacy in schools with higher SWPBS implementation scores (Ross et al., 2012). Conversely, Oakes, Lane, Jenkins, and Booker (2013) observed no differences in perceived self-efficacy between teachers or between groups of high- and low-implementing teachers in a study of two middle schools that had implemented a model called "Comprehensive, Integrated, Three-Tiered Model of Prevention" for 1 year. This model was an extended version of the original SWPBS model, in which academic, behavioral, and social skills programming were addressed within one integrated framework (Oakes et al., 2013). Also, in a randomized study, Bradshaw, Pas, Goldweber, Rosenberg, and Leaf (2012) compared teacher self-efficacy in 22 schools implementing the standard SWPBS model with 20 schools implementing an adapted version called PBIS*plus*. In PBIS*plus*, implementation of tier two supports is specifically addressed, and the schools are provided tailored training by an external coach in functional behavioral assessments, student support teaming process, cultural adaptation, and evidence-based practices. A significant intervention effect on teacher self-efficacy favoring the adapted version was identified (Bradshaw et al., 2012).

Thus, prior research indicates (a) a consistent, moderate, and most likely reciprocal association between teachers' efficacy beliefs and student outcomes, both academically and behaviorally and (b) that the SWPBS model may positively affect the teachers' individual sense of efficacy, particularly in high-implementation schools. We hypothesized that the model as implemented in Norway would have positive influences on staffs' self-efficacy beliefs as well as on their perceptions of the capability of the entire collegium.

Behavior Management and Student Behavior

Improving teachers' use of evidence-based, proactive, positive practices in their handling of misconduct and vulnerable students is another "cornerstone" of SWPBS. Teachers' behavior management is assumed to influence student behavior primarily by reinforcement. Many teachers, however, appear unaware of their influence on student behavior and of the reciprocal relationship between student behavior and the teacher's own behavior (Sutherland & Oswald, 2005). Student problem behavior is partially fostered by ineffective instruction and punitive disciplinary practices, which often characterize the school system's response to problem behavior (e.g., Clunies-Ross, Little, & Kienhuis, 2008).

Behavior management comprises both positive behavior support and behavioral corrections and is sometimes referred to as classroom management. *Positive behavior support*

addresses proactive strategies that influence student behavior by teaching and learning the rules and norms of expected behavior and by supporting prosocial behavior such as praise, positive attention, pre-correction, active behavior supervision, and rewards (token economy). Effective *behavioral corrections* include predictable responses to inappropriate behavior in the form of mild consequences such as reprimands, re-learning of rules/expected behavior, time-outs, or parent reports.

There is no empirical evidence that harsh and intrusive reactions to student problem behavior such as suspension, expulsion, detention or referring students to the principal have any beneficial effects or deter future misconduct (e.g., American Academy of Pediatrics Council on School Health, 2013). Conversely, mild and less intrusive strategies such as verbal reprimands, commands, and removal of reinforcing activities following rule-breaking behavior may have positive effects on disruptive classroom behavior (e.g., Nelson & Roberts, 2000). Such responses appear more effective if combined with contingent reinforcement of prosocial behaviors, as recommended in the SWPBS model (Sprague & Walker, 2005).

Contingent use of positive social reinforcement has repeatedly been shown to have positive effects on student conduct (e.g. Oliver et al., 2011) and an effective classroom management tool (e.g., Maggin, Chafouleas, Goddard, & Johnson, 2011). Moreover, a few, clear, and positively stated rules have been found to increase students' on-task behavior and willingness to follow school rules (e.g., Gable, Hester, Rock, & Hughes, 2009). As emphasized in SWPBS, effective behavior management, including adequate and systematic responses to disruptive behavior, appears to be a prerequisite for efficient handling students at moderate to high risk of conduct disorder.

Effects of SWPBS on Behavior Management in School

With the exception of two single-school studies, we have not been able to identify relevant prior studies investigating the impacts of SWPBS on the teacher behavior. In their study, DePry and Sugai (2002) followed one sixth-grade teacher in a rural SWPBS school for a month. The focus was on immediate effects of the teacher's behavior management practices on minor behavioral incidents in class. The data indicated a positive relation between teacher behavior and student behavior: a relatively low level of active supervision and pre-correction resulted in moderate decreases in the students' problem behavior. Myers, Simonsen, and Sugai (2011) conducted their multiple baseline study among four teachers in a middle school (Grades 5-8) that was in its second year of SWPBS implementation. As hypothesized, the study results indicated that teachers increased their rates of praise statements when provided with a continuum of performance feedback dependent on their individual needs. Parallel to the increase in use of praise, a downward trend in student off-task and disruptive behavior was observed.

Taken together, prior studies indicate positive teacher outcomes SWPBS, although to date, no valid conclusions regarding intervention effects can be drawn. The obvious gap in knowledge regarding teacher outcomes in general and the effects on collective efficacy and behavior management in particular as well as the lack of larger scale studies with rigorous evaluation designs inspired a focus on the impacts of N-PALS on staff behavior and perceived efficacy in the present study.

SWPBS in Norway

The N-PALS is an culturally adapted version of the widely disseminated version of the U.S. SWPBS model developed by Sprague and Walker (2005) and implemented in close collaboration with the program developers at the University of Oregon (Sprague & Walker, 2005). The Norwegian adaptations were mostly surface changes while the following core components remain the same in the two countries: (a) an intervention program with pre-defined components targeting all students, (b) a training program for the staff, and (c) school-wide strategies for the implementation and sustainability of change. One remarkable difference is the Norwegian emphasis on developing a nationwide rather than a community or statewide standardized system of implementation and quality assurance. Although structured, SWPBS/N-PALS is a flexible system-level approach to the prevention of student problem behavior and to the promotion of a positive and inclusive school environment that can facilitate teaching, optimal learning, and psycho-social functioning for all students (e.g., Arnesen, Meek-Hansen, Ogden, & Sørlie, 2014). The focus is on positive, systematic, supportive, collective (school-wide), proactive, and inclusive practices. N-PALS builds on the principle of inclusion, systemic approaches, and competence building at the school level; positive behavior support for all students with interventions at three levels; and assessment, review, and follow-up activities on all three levels. Punishing approaches to problem behavior are replaced with proactive strategies which stress the importance of positive behavior support and consistency in the communication of norms and rules in school. Student problem behavior is prevented and reduced through teaching and learning positively formulated rules, consistent support for positive behavior, and predictable consequences for unwanted or negative behavior. A continuum of evidence-based interventions is targeted to the students' risk level and the schools' needs in a three-tiered model organized according to the principle of "response to intervention" (RtI). As a result, a safe and positive learning climate is established for all students at the universal level, while social or behavioral problems among individual students are targeted at the selected and indicated levels (Arnesen et al., 2014).

The *universal* level builds on the assumption that as many as 80% to 90% of the students are well behaved, but still they deserve praise and encouragement for complying with school expectations. Interventions at this level include (a) teaching school-wide rules and consistent praise and encouragement of positive student behavior; (b) applying predictable, but mild consequences for problem behavior; (c) training teachers in proactive classroom management and good directions; and (d) school-wide monitoring of student behavior using the School-Wide-Information system (SWIS). The 5% to 15% of the students who do not respond satisfactorily to the universal interventions is targeted at the *selected* level. This group of students may receive time-limited small group training in academic or social skills or they are offered the Check-In/Check-Out (CICO) program (Todd, Campbell, Meyer, & Horner, 2008), which efficiently addresses the students' needs for intensive and targeted support. Individually tailored interventions are organized at the *indicated* level for students at high risk of conduct disorder and school failure (1%-5%). Interventions at this level may include (a) individual and multisystemic support plans based on functional behavioral assessment (FBA), (b) individual social skills training (e.g., "Stop-Now-and-Plan"; Augimeri, Farrington, Koegl, & Day, 2007), (c) parent training, and (d) behavior management training for teachers.

Implementation Structure

In Norway, the NCCBD has the sole national responsibility for the implementation and sustainability of N-PALS, including quality assurance procedures. The three-tiered model, the training of coaches and the internal school implementation teams, web-based information systems for regular measurement of implementation quality, and student behavior, are nationally standardized. The implementation process at each school takes about 3 to 5 years, and starts with a formal assessment of the schools' readiness for implementation, and approval from a majority of the staff (min. 80%) is required, along with active engagement from the school leadership and parental support. The school is also required to define at least one goal of improvement, commit to prioritize the program for 3 to 5 years, and contribute with monitoring and evaluation of the implementation and outcomes.

At each school a team is established with five to seven participants representing school management, teachers, other staff, school psychological services, and parents. A student representative may also be appointed. The school teams are trained and supervised by local coaches (40 hr over 2 years) who are trained for 1 year and certified by the program implementers at the NCCBD. In addition, regional network meetings and booster sessions are provided quarterly. The school team leads and coordinates the activities at the school, including the implementation and evaluation of the school's action plan which is based on a school-wide assessment of strengths and challenges. The teams also are responsible for continuous data-driven analyses for interventions and sustained implementation fidelity. They keep close contact with the school staff and organize problem-solving discussions and training in key elements of the model and implementation structure.

Research Questions

The research questions were as follows: (a) Do staff in schools that implement the N-PALS model over a 3-year period demonstrate higher perceived efficacy (self-efficacy and collective efficacy) across time than staff in control schools doing "practice-as-usual"? (b) Do the staff in N-PALS and control schools differ in behavior management (i.e., use of positive behavior support and behavioral correction) as observed by students and staff? and (c) Are intervention outcomes moderated by implementation quality?

Method

The effectiveness study was based on a quasi-experimental matched comparison group design (for a detailed description, see Sørlie & Ogden, 2014). The presented results are based on data collected from all staff members (principals, teachers, assistants, special education teachers, after-school personnel) and students in fourth to seventh grades in 28 N-PALS schools and 20 comparison schools doing "regular practice." The staff and students completed questionnaires at four time points (T1-T4) over a 4-year period. The study had an open cohort design, allowing new staff and students to be enrolled at each measure point. Participants who left the schools were not followed due to a lack of resources. Pre-test (T1) was at the end of the school year prior to N-PALS. T2 was at the beginning of a new school year and close to the initiation of the intervention (6 months past T1). The students participated from T2 forward. T3 and T4 (post-test) were conducted after, respectively, 1 and 3 years of implementation.

Participants

The intervention and control groups were compared at baseline on several school, student, and staff variables such as school size, the proportion of students with immigrant backgrounds, proportion of students in special education, academic test scores, staff gender, and educational background. Significant differences were identified in one of the 29 variables tested (examples below). The baseline analyses indicated that the two school groups were comparable, which is a prerequisite for valid estimates of intervention effects. Possible group differences on unobserved variables can, however, not be ruled out. In the 48 participating schools (Grades 1-7), the student body numbered 13,570 at baseline (N-PALS = 7,964, control = 5,606), of whom 51% were boys and approximately 6% had ethnic minority backgrounds. No

initial differences appeared between the intervention and comparison group with the exception of reading performance (test scores in fifth grade). Somewhat lower test scores were observed in the N-PALS group than in the comparison group, $F(1, 47) = 4.81$, $p = .03$. The participating schools matched the national average both on school size, standardized test performance scores, and student ratings of the learning environment (Statistics Norway, 2009). At T1, the total staff numbered 1,064 in the intervention schools and 750 in the comparison school. However, only staff members in daily contact with groups of students were invited as respondents ($N = 1,266$). Sixty-four percent of the school staff was teachers, 16.5% were assistants, and 22% were after-school personnel. Most were trained as teachers (56%), 12% had additional special education training, while 4% had no kind of formal training. About 80% were females and older than 35 years. No initial group differences in staff characteristics were identified.

Procedures

To match the comparison group with the intervention group on key characteristics, all primary schools in 17 municipalities (in southern and western Norway) were matched on schools size (>100 students) and geographical location prior to invitation. Schools actively implementing structured programs other than N-PALS were excluded. Of the 126 schools considered eligible for the study, 52 were randomly selected and invited as intervention schools. Of these, 28 (54%) agreed to participate. Next, another group of schools ($n = 44$) were randomly selected and invited to participate in a longitudinal study of student behavior and informed that they would serve as comparison schools. Only schools located more than five kilometers away from an intervention school were invited, this to minimize program contamination. Of these, 20 schools agreed to participate (45.5%). No significant differences were found between the participating schools and the declining schools. Questionnaires were completed during ordinary working time or class time. Written instructions were given to standardize the assessment procedures. Informed and written consent from parents was obtained in advance, while agreement from staff was obtained coincident with completing the questionnaire.

Measures

Collective and self-efficacy in school. Perceived collective efficacy in school was measured with a frequently used 12-item scale developed by Goddard (2002), a revised version of the well-validated Collective Efficacy Scale (CES; Goddard, Hoy, & Woolfolk Hoy, 2000). CES assess the extent to which teachers in a school believe in their mutual capability to positively influence the students' learning and school outcomes. Items were rated by staff on a 5-point scale (1 = *never*, 5 = *very often*) and included items such as "teachers here are confident they will be able to motivate their students" and "teachers in this school are able to get through to difficult students." The internal consistency in the current study was acceptable with Cronbach's alpha (α) ranging from .82 to .85 across assessment points. Perceived *s*elf-efficacy in school was measured by a 30-item scale developed for the study. Each staff member was requested to rate how competent (skillful) she or he felt in managing and preventing problem behavior and promoting the students' academic skills. A 7-point rating scale (1 = *highly incapable*, 7 = *highly capable*) was used on items such as "to stop aggressive student behavior" and "be present in the classroom when the lesson begins." Cronbach's alpha was .96 at all four assessment points.

Behavior management. Strategies to promote positive student behavior and to manage student problem behavior were assessed by both students and staff. A 17-item staff scale and an equivalent 23-item student scale were developed for this study. Factor analyses revealed a two-factor structure: "positive behavior support" (staff: nine items, α from .74 to .76; students: 16 items, α from .88 to .92) and "behavioral correction" (staff: eight items, α from .57 to .66; students: five items, α from .58 to .61). The staff reported the number of times in the last month they had used strategies such as "praised and encouraged expected positive student behavior," "deliberately ignored undesirable behavior." The students reported how often in the last month they had experienced teacher practices such as "the teacher praises me when I do as she/he says during lessons," and "the teacher says what will happen if we violate the rules." A 7-point scoring scale was used in the staff version, ranging from 0 (*0 times*) to 7 (*more than 20 times*) whereas the student version used a 4-point scale, ranging from 1 (*never*) to 4 (*usually*). We hypothesized that to reduce the amount of student problem behavior, the ratio between positive supports and behavioral corrections is essential. A rule of thumb in SWPBS/N-PALS is that teachers should encourage and reward the students' positive behavior three to five times for every behavioral correction.

Implementation quality. To assess the implementation fidelity, the "PBIS Self-Assessment Survey" (SAS, 46 items) was completed by school staff in the intervention schools (Sugai, Horner, & Todd, 2009). SAS has shown satisfactory psychometric qualities in several prior evaluation studies and measures fidelity at respectively the school-wide, classroom, non-classroom, and student level. In the present study, the staff members rated how statements corresponded with the actual situation at their school, using a 3-point scale (1 = *in place*, 2 = *partially in place*, 3 = *not in place*). Examples of statements are "a few (3-5) school rules are positively and clearly defined" (school-wide), "problem behavior is consistently addressed with mild and predictable negative consequences (1:5)" (classroom), "the staff members who have inspection actively supervise the students in all arenas outside the classroom context" (common areas), and "a

behavioral support team is immediately established (within 2 days) for students showing severe behavior problems" (individual). The alphas for the sumscale and subscales ranged from .90 to .92 across assessment points.

Because contextual factors may influence both intervention outcomes and implementation quality, several school factors were included as covariates in the analyses of N-PALS effects: school size, proportion of students with immigrant backgrounds, proportion of special education students, proportion of staff without any formal training, and proportion of staff working in after-school services.

Analytic Approach

Multilevel analyses with three levels were conducted to compare change across time in the control and intervention group. The longitudinal data structure, with repeated measurements as Level 1 (T1, T2, T3, T4) nested within staff or students at Level 2 and schools at Level 3, suggested a multilevel approach. The Mixed Linear procedure in IBM SPSS statistics (version 20) was used to account for the hierarchical structures in the data. In the main effect analyses, the control group was used as reference group, baseline was used as the reference time point, and T2 to T4 were included in the analyses. In addition, contextual school variables that could potentially influence the effects were statistically controlled for. All covariates were centered to minimize collinearity, and non-normally distributed variables were log-transformed. Because of the differences in time between measurements, an unstructured residual covariance structure was chosen for residuals at Level 1. At the school level, random intercepts were estimated using a scaled identity covariance structure. Main effects were investigated by including a Time × Group interaction term in the models. Moderating effects of implementation quality in the N-PALS schools were examined in a second series of models. These models included the same variables as the main effect models, but were extended by including a two-way interaction term, Time × Implementation Quality. All relevant lower-order terms were included to ensure balanced regressions equations. Separate analyses were run for the teacher and student rated outcome variables. Missing data were estimated using the direct-likelihood method, assuming a missing-at-random mechanism. We calculated Cohen's *d* based on Feingold's (2013) recommendations for designs with two independent groups and linear models with three or more time points. Confidence intervals were calculated based on the procedure suggested by Grissom and Kim (2005). Unconditional variable descriptives are shown in Table 1.

Results

Attrition

At the first assessment, 75% of the staff members participated; 76% participated at T2; 73% participated at T3; and 67% participated at post-test. Of the students, 94% with parental consent participated at T2 and T3, and 91.9% participated at T4. Throughout the study period, a total of 11,784 students in fourth to seventh grade were asked to participate. Of these, 8,474 students got parental consent (72%). Variations in participation among staff were primarily due to turnover ($n = 560$) and leaves of absence during the data collection periods ($n = 423$). The open cohort design allowing inclusion of new fourth graders and other students new to the schools at each time point, and older students leaving the study by the end of seventh grade accounted for most of the missing data and varying student participation rates. In addition, some missing data were related to students changing schools ($n = 360$), parents withdrawing their consent ($n = 44$), or absence on the assessment day. Furthermore, one of the control school withdrew from the study prior to T2 ($n_{staff} = 59$) and one before T3 ($n_{staff} = 30$, $n_{students} = 221$), stating work overload as the main reason. Missing on the outcome variables by time point is shown in Table 2 (varied from 0.8% to 3.6%). School-level variables did not have missing.

Baseline comparisons of staff present and missing at post-test showed that those not present were more likely to have a university degree, $F(1, 1209) = 10.591$, $p < .001$, and to be working in higher rather than lower grades, $F(1, 1209) = 8.565$, $p < .01$. Comparisons between students participating and those missing at post-test showed that the missing group rated their learning climate slightly more positive at baseline, $F(1, 2427) = 14.76$, $p < .001$. No school group or gender differences were found in the staff and student analyses. In conclusion, the attrition was rather high, but not higher than could be expected from the open cohort design and normal fluctuations in the staff and student populations across time. Together with few significant differences between those who participated at pre- and post-assessment and those who did not, the results suggest that a missing-at-random mechanism may be assumed, and that estimation of missing data by the direct-likelihood method is acceptable.

Main Intervention Effects

Indications of positive 3-year main effects of the N-PALS model were found for collective efficacy, self-efficacy, and positive behavior support practices as rated by the school staffs (see Table 3, Figure 1). To simplify the often rather difficult-to-read results of multilevel analysis, the estimates of fixed effects in Table 3 are shown as group differences in change across measure points. The table also includes effect sizes (Cohen's *d*) with 95% confidence intervals. The intervention effects ranged from small to large with Cohen's *d* between .14 and .91.

Perceived teacher efficacy. As shown in Table 3, there was positive growth across time in the school staffs' collective efficacy scores in both school groups. The multilevel analysis showed that the development in the N-PALS group did

Table 1. Unconditional Means and Standard Deviations by Group and Time.

	Intervention group					Control group				
Variable	T1	T2	T3	T4	T5	T1	T2	T3	T4	T5
Collective efficacy	55.03 (7.01)	55.89 (6.92)	56.28 (7.25)	—	58.24 (6.39)	57.80 (6.47)	58.48 (6.48)	58.75 (6.50)	—	59.09 (6.43)
Self-efficacy	148.99 (17.48)	148.65 (17.58)	151.18 (19.01)	—	155.41 (18.43)	155.63 (18.35)	154.53 (19.39)	157.11 (18.80)	—	160.85 (18.88)
Positive behavior support										
Staff	13.65 (7.42)	13.69 (7.43)	13.96 (7.35)	—	20.78 (7.43)	13.11 (6.80)	12.55 (6.61)	12.32 (6.86)	—	13.61 (7.53)
Student	—	53.89 (7.36)	52.38 (8.80)	53.11 (8.30)	53.46 (7.87)	—	54.65 (6.79)	52.73 (8.27)	53.00 (8.12)	53.42 (7.87)
Behavioral correction										
Staff	4.87 (3.24)	4.62 (3.26)	4.29 (2.93)	—	4.15 (3.24)	4.63 (3.10)	4.11 (3.03)	4.08 (3.32)	—	3.91 (3.02)
Student	—	16.27 (2.57)	16.12 (2.78)	16.18 (2.69)	16.22 (2.65)	—	16.53 (2.46)	16.14 (2.74)	16.26 (2.62)	16.38 (2.65)
Portion immigrant background	5.56 (6.24)	5.00 (5.16)	5.35 (5.41)	—	5.20 (5.23)	6.89 (4.97)	6.77 (5.27)	5.62 (3.71)	—	5.78 (3.82)
Portion special education	5.24 (3.88)	5.17 (3.37)	5.36 (3.36)	—	5.10 (3.27)	4.74 (2.70)	5.04 (2.75)	4.85 (2.86)	—	4.79 (2.76)
Portion unqualified staff	4.05 (7.51)	3.70 (7.13)	3.68 (7.14)	—	3.60 (7.03)	2.79 (6.32)	3.20 (7.10)	3.69 (8.23)	—	2.64 (5.81)
Implementation quality	—	—	87.36 (3.93)	106.63 (6.55)	116.06 (6.57)	—	—	—	—	—

Note. — = not measured.

Table 2. Missing Data by Variable and Time Point (Percentage).

Variable	T1	T2	T3	T4
Perceived collective efficacy	0.9	1.8	1.9	1.7
Perceived self-efficacy	1.7	3.0	2.4	2.8
Positive behavior support (student)	3.6	3.3	1.9	2.4
Positive behavior support (teacher)	0.9	1.4	1.6	2.4
Behavioral correction (student)	1.7	1.7	0.9	1.3
Behavioral correction (teacher)	0.8	1.3	1.5	2.4

not differ significantly from the development in the control group, neither prior to intervention (pre-T2 diff = .11, p = .756, not in table) nor after 1 year of implementation (pre-Y1 diff = .57, p = 756). However, after 3 years of implementation, significantly higher perceived collective efficacy over time was observed in the intervention group as compared with the control (pre–post diff = 2.28, p = .000). Similar progress was observed for self-efficacy, indicating a significant main effect of N-PALS (pre–post diff = 2. 53, p = .014).

Behavior management. A main intervention effect was also observed in disciplinary practices: The staff in the N-PALS schools increased the amount of proactive and supportive practices considerably more from pre- to post-assessment than did their colleagues in the comparison schools (pre–post diff = 6.56, p = .000). Although there was a general decrease in the use of behavioral corrections across time according to the staff ratings, no significant group differences were observed. The student ratings of the teachers' behavior management practices indicated no significant group differences in change across time. Additional analyses revealed no impacts either by gender or class level on the students' ratings. However, according to the staff ratings and in line with the recommended "rule of thumb" (3-5:1, Arnesen et al., 2014), the ratio between positive supports and corrections in the N-PALS group increased from 2.5 at pretest to 4.5 at post-test (control: 2.5 - 3.1), while remaining stable at 1.03 in both groups according to the students.

Implementation Quality

Twenty-six of the 28 intervention schools reached a total implementation score in the range of 75% to 90% after 3 years, while two schools reached respectively 72% and 73%. Differential effects of N-PALS were explored by the running of moderation analyses of fidelity while other possibly influencing factors were accounted for (school, staff, and student characteristics). No significant moderation effect was identified in any of the outcome variables, which may be a consequence of scant variation in implementation scores among the intervention schools.

Discussion

In the present article, we examined the impacts of N-PALS on school staffs' practices and on their individual and collective efficacy in a strengthened, non-randomized experiment with 28 intervention schools and 20 "regular practice" schools serving as controls. We also examined whether the quality of the implementation moderated the outcomes.

Results after 3 years of implementation indicated that this school-wide model had significant positive effects according to school staff. Consistent with prior studies on the SWPBS model in the United States and Canada, a significant main effect was observed in the N-PALS staffs' perceived *self-efficacy* compared with their colleagues in the control group. This finding suggests that individual staff members felt more empowered, more motivated, and more capable of teaching and handling challenging students following the implementation of N-PALS.

The study results also indicated a positive main effect on school staffs' *collective efficacy* beliefs compared with pre–post changes in the control group. The finding is of interest because prior research has found that collective efficacy is a stable school-level characteristic that varies between schools (Sørlie & Torsheim, 2011) and is significantly related both to the students' academic performance and to the amount of problem behavior in schools (Sørlie & Torsheim, 2011).

Moreover, the results indicated a considerable main effect of N-PALS on staffs' behavior management practices as rated by the staff members themselves. The use of positive behavior-supporting strategies such as verbal praise for prosocial behavior increased from pre- to post-test in both groups, but substantially more so in the intervention schools than in the control schools (d = .91). The ratings of behavioral corrections such as verbal reprimands and mild sanctions for unacceptable school behavior showed a decreasing trend in both conditions; however, no group differences in change emerged. However, the ratio of supporting to corrective reactions increased more from pre- to post-test in the N-PALS schools than in the comparison schools. In other words, the staffs in the interventions schools simultaneously increased their positive feedback and decreased their negative feedback to students.

These findings are important given that research indicates that school staffs with poor behavior management skills contribute to placing students at risk for higher levels of problem behavior, poor academic outcomes over a long period of time, and exclusion from ordinary education (e.g., Reinke & Herman, 2002). Poor behavior management skills are also related to negative student–teacher relationships, off-task behavior, and poor learning environments because of frequent classroom disruptions (e.g., Korpershoek, Harms, de Boer, van Kuijk, & Doolaard, 2014). Authoritative classroom management and efficacious prevention and management of problem behavior require that teachers increase their positive behavior support at the same time as they reduce the amount of harsh discipline they impose (e.g., Reinke, Lewis-Palmer, & Merrell, 2008).

Table 3. Main Effects of N-PALS: Differences in Change Across Time Within and Between the Intervention and Control Group—Fixed Effects Estimates and Effect Sizes (Cohen's d) With Lower and Upper Confidence Intervals.

	Change within group									
	N-PALS		Control		Change between groups			Effect Size (ES)		
	Pre-Y1	Pre–post	Pre-Y1	Pre–post	Pre-Y1	*p*	Pre–post	*p*	*d*	*d* (95% CIΔ)
Collective efficacy	1.08	3.13	0.71	0.84	0.57	.756	2.28	.000	0.34	[0.336, 0.344]
Self-efficacy	2.15	6.54	0.87	4.01	1.28	.176	2.53	.014	0.14	[0.126, 0.144]
Behavior management										
Positive behavior	−0.17	6.72	−0.88	0.16	0.71	.133	6.56	.000	0.91	[0.906, 0.914]
support[a]	1.71	0.34	2.16	0.24	−0.46	.108	0.10	.709	0.01	[0.009, 0.011]
Behavioral correction[a]	−0.75	−0.93	−0.67	−0.92	−0.09	.642	−0.02	.943	0.01	[0.006, 0.014]
	0.21	0.05	0.24	−0.09	−0.03	.760	0.13	.167	0.05	[0.049, 0.051]

Note. Estimates (enhanced values) are based on Satterthwaite's (1947) approximate degrees of freedom. All covariates (total number of students, portion of students with foreign background, portion of special education students, and portion on staff without formal education, portion of staff working in after-school services) and the Time × Group interaction are accounted for in the estimates of change. Model estimates for all variables in the outcome analyses are available on request to corresponding author. N-PALS = Norwegian version of SWPBS; pre-Y1 = change from first measure point for the staff (T1) and students (T2) to the end of year one with N-PALS; pre–post = change from first measure point for staff (T1) and students (T2) to the end of year three with N-PALS; CI = confidence interval; SWPBS = School-Wide Positive Behavior Support.
[a]Staff ratings on first line, student ratings on second line.

Greater positive changes across time in individual and collective efficacy staff-reported behavior management in the N-PALS schools may stem from a combination of (a) the staffs' observations of reduced student behavior problems during implementation of the model (Sørlie & Ogden, 2015) in accordance with the reciprocal relationship between the phenomena; (b) the model's strong focus on a collective (school-wide) approach and mutual efforts to develop a more positive, supportive, and inclusive school environment; and (c) more systematic and evidence-based practices. It is also reasonable to assume that the reported changes may partially stem from the in-school training in behavior management.

Contrary to what was expected, no significant effects of the intervention were observed in the students' (Grades 4 to 7) assessments of the teachers' disciplinary practices. Although teachers are seen as highly reliable raters of their own practices (Desimone, Smith, & Frisvold, 2010), we cannot rule out that the discrepancy between staff and student ratings may partially be due to positive bias in the staff ratings. The discrepancy may also reflect that the changes in the teachers' practices were too slight to be noticed by the students due to too low classroom-level fidelity of implementation. Or perhaps the students were not sufficiently sensitive to changes in the *amount* of behavioral feedback they received. We speculate that the students primarily perceived that they were receiving some response from teachers, and how the attention made them feel. Accordingly, the validity of student ratings (9-12 years) may be questioned with regard to the number of positive and negative behavioral responses they received over a 30-days period, particularly when rated retrospectively, as in the current study (i.e., "last month"). Besides, delayed effects may be part of the explanation. Delayed effects relate to that it may take time to induce changes in teachers' behavior and that intervention effects thus occur after an extended period of time (i.e., the study time frame may have been insufficient). Moreover, modest between-rater agreement is a familiar phenomenon, and similar informant discrepancies have been reported in prior intervention trials (De Los Reyes, 2011).

Although schools implementing other school-wide programs were not invited to the study (exclusion criteria), it should be noted that all but three of the control schools reported that they had been implementing at least one evidence-based program *during* the study period (e.g., bullying prevention, social skills promotion). Thus, N-PALS was evaluated using a harder test than initially intended, and the outcomes may be considered conservative estimates of intervention effects.

Differential Effects

Among the moderators examined in this kind of studies, high implementation quality is often found to predict positive outcomes across settings and programs (e.g., Ross et al., 2012), while small or no effects often relate to low fidelity scores. In this study, however, implementation quality did not moderate any of the outcomes. Limited variations in the implementation quality among N-PALS schools were found, evidenced by the finding that at posttest 80% of the schools had reached the recommended SAS threshold score. This could be taken to mean that the intervention model was consistently implemented across schools and that the intervention schools benefitted approximately equally from the model.

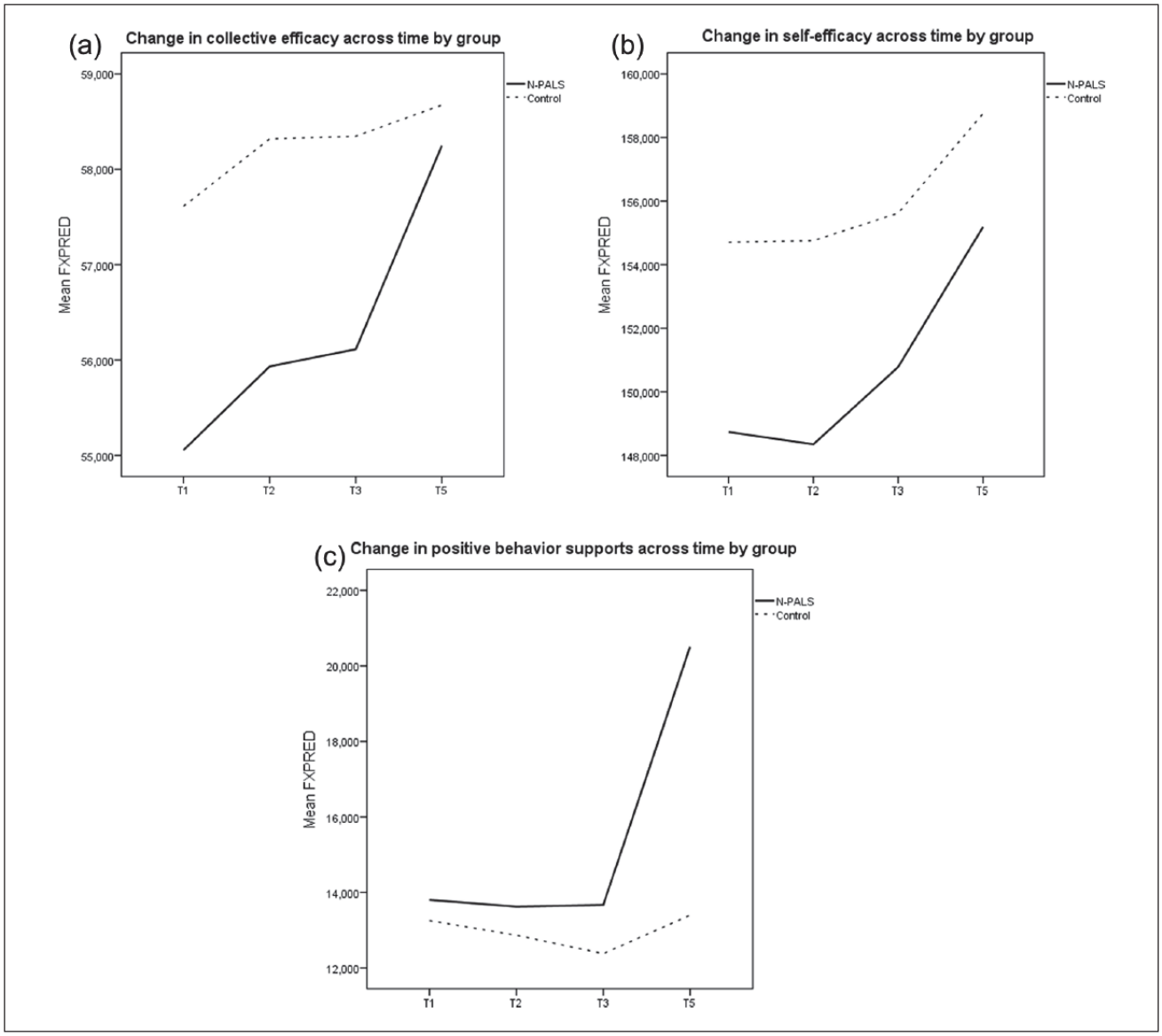


Figure 1. Main intervention effects: Changes across time in the N-PALS and control group.
Note. N-PALS = Norwegian version of SWPBS; SWPBS = School-Wide Positive Behavior Support.

Limitations

Among the limitations of the study, the first concerns selection bias in the process of recruiting schools to the study. Even if the intervention and comparison group where matched on key variables, undetected group differences in non-observed variables may occur. It should also be mentioned that not all of the invited intervention and control schools accepted the invitation. But we assume that any selection bias caused by refusals would be equally large in both groups, and in support of this assumption, no differences were found among the declining and accepting school groups.

A second limitation could be caused by the fact that the staff members in the N-PALS schools were the primary implementers of the intervention model, and consequently a positive response bias may have occurred in their assessments of the outcomes. Although we cannot rule out this allegiance effect, a recent meta-analysis revealed that teachers' self-reports on teaching are highly reliable, showing strong correlations with both classroom observations and teachers' records (Desimone et al., 2010). We also find the staff assessments highly relevant in this context, because they are the primary stakeholders in efforts to increase teacher competence and positive student behavior. A third limitation is the low correlations between the staff- and

student-based assessments. Such discrepancies between different informants are common, and the causes of these discrepancies remain unclear. There is a concern in the present study that some of the assessment instruments may have been insufficiently matched to the developmental level of the students, and that direct observations in the classrooms or student interviews could have added to the validity of the study.

A possible fourth limitation relates to SAS as the sole implementation measure.

Although several studies have used SAS as an implementation measure, it was originally developed for needs assessment, not specifically as a fidelity measure. Additional, however, more resource-demanding measures developed for measuring fidelity in SWPBS schools could have been used, but in the present study this was considered too expensive and time-consuming.

Implications for Future Research

Future replications of this study should preferably apply a randomized controlled design and add observational data of teacher practice to strengthen the conclusions regarding the N-PALS' impacts for school staffs. Future efforts should also include analyses of relevant moderators other than implementation quality (e.g., school size) and attempt to reveal model components and implementation items that may improve the effectiveness of N-PALS. Investigating potential long-term effects of the model is also called for, and will be the topic of a new study beginning in 2015 using national school registry data.

Conclusions

The SWPBS model as implemented in Norway appeared effective in promoting school staffs' perceived individual and collective efficacy. The model also appeared effective in changing staffs' behavior in directions that are more positive and demonstrably more efficient to prevent and address student problem behavior, at least in staffs' own eyes. The staff members in the N-PALS schools substantially increased their use of positive behavior supports whereas a co-occurring decrease (even non-significant) in behavioral corrections was observed. However, the effects observed in the staff ratings did not seem to translate to the students' ratings, which need further research. The Norwegian outcomes indicate that the SWPBS model may be successfully transferred across the Atlantic without major adaptations, at least to Norway as a representative of the Scandinavian and western European countries.

Authors' Note

The standards of the Regional Committee for Medical and Health Research Ethics (REK South-East) and of the Norwegian Social Science Data Services were followed throughout the conduct of the study.

Acknowledgments

We acknowledge Professor Torbjørn Torsheim for his invaluable support with the analyses. We are also grateful for the intervention adapters', Anne Arnesen and Wilhelm Meek-Hansen, genuine interest in the study. None of them has been involved in the research.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research and/or authorship of this article: This research was supported by the Norwegian Center for Child Behavioral Research and in part by a grant from the Norwegian Directorate for Education and Training.

References

Almog, O., & Shechtman, Z. (2007). Teachers' democratic and efficacy beliefs and styles of coping with behavioral problems of pupils with special needs. *European Journal of Special Needs Education*, *2*, 115-129.

American Academy of Pediatrics Council on School Health. (2013). Policy statement: Out-of-school suspension and expulsion. *Pediatrics*, *131*, e1000-e1007. doi:10.1542/peds.2012-3932

Arnesen, A., Meek-Hansen, W., Ogden, T., & Sørlie, M.-A. (2014). *Positiv læringsstøtte i skolen. Hele skolen med* [Positive learning support in school. The whole school participates]. Oslo, Norway: Universitetsforlaget.

Augimeri, L. K., Farrington, D. P., Koegl, C. J., & Day, D. M. (2007). The SNAP™ under 12 outreach project: Effects of a community based program for children with conduct problems. *Journal of Child and Family Studies*, *16*, 799-807.

Bradshaw, C. P., Pas, E. T., Goldweber, A., Rosenberg, M. S., & Leaf, P. J. (2012). Integrating school-wide positive behavioral interventions and supports with tier 2 coaching to student support teams: The PBISplus model. *Advances in School Mental Health Promotion*, *5*, 177-193.

Brouwers, A., & Tomic, W. (2000). A longitudinal study of teacher burnout and perceived self-efficacy in classroom management. *Teaching and Teacher Education*, *16*, 239-253.

Clunies-Ross, P., Little, E., & Kienhuis, M. (2008). Self-reported and actual use of proactive and reactive classroom management strategies and their relationship with teacher stress and student behavior. *Educational Psychology: An International Journal of Experimental Educational Psychology*, *28*, 693-710.

De Los Reyes, A. (2011). Introduction to the special section: More than measurement error: Discovering meaning behind informant discrepancies in clinical assessments of children and adolescents. *Journal of Clinical Child & Adolescent Psychology*, *40*, 1-9.

DePry, R. L., & Sugai, G. (2002). The effect of active supervision and pre-correction on minor behavioral incidents in a sixth

grade general education classroom. *Journal of Behavioral Education*, *11*, 255-267.

Desimone, L., Smith, T., & Frisvold, D. (2010). Survey measures of classroom instruction: Comparing student and teacher reports. *Educational Policy*, *24*, 267-329.

Feingold, A. (2013). A regression framework for effect size assessments in longitudinal modeling of group differences. *Review of General Psychology*, *1*, 111-121.

Gable, R. A., Hester, P. H., Rock, M. L., & Hughes, K. G. (2009). Back to basics: Rules, praise, ignoring, and reprimands revisited. *Intervention in School and Clinic*, *44*, 195-205.

Gibbs, S., & Powell, B. (2011). Teacher efficacy and pupil behavior: The structure of teachers' individual and collective beliefs and their relationship with numbers of pupils excluded from school. *British Journal of Educational Psychology*, *82*, 564-584.

Goddard, R. D. (2002). A theoretical and empirical analysis of the measurement of collective efficacy: The development of a short form. *Educational and Psychological Measurement*, *62*, 97-110.

Goddard, R. D., Hoy, W. K., & Woolfolk Hoy, A. (2000). Collective teacher efficacy: Its meaning, measure, and impact on student achievement. *American Educational Research Journal*, *37*, 479-507.

Goddard, R. D., Hoy, W. K., & Woolfolk Hoy, A. (2004). Collective efficacy beliefs: Theoretical developments, empirical evidence, and future directions. *Educational Researcher*, *33*, 3-13.

Grissom, R. J., & Kim, J. J. (2005). *Effect sizes for research: A broad practical approach*. Mahwah, NJ: Lawrence Erlbaum.

Jordan, A., & Stanovich, P. (2003). Teachers' personal epistemological beliefs about students with disabilities as indicators of effective teaching practices. *Journal of Research in Special Educational Needs*, *1*, 1-2.

Kelm, J. L., & McIntosh, K. (2012). Effects of School-Wide Positive Behavior Support on teacher self-efficacy. *Psychology in the School*, *49*, 137-147.

Klassen, R. M., & Tze, V. M. (2014). Teachers' self-efficacy, personality, and teaching effectiveness: A meta-analysis. *Educational Research Review*, *12*, 59-76.

Korpershoek, H., Harms, T., de Boer, H., van Kuijk, M., & Doolaard, S. (2014). *Effective classroom management strategies and classroom management programs for educational practice*. Groningen, The Netherland: GION onderwijs/onderzoek, Rijksuniversiteit.

Maggin, D. M., Chafouleas, S. M., Goddard, K. M., & Johnson, A. H. (2011). A systematic evaluation of token economies as a classroom management tool for students with challenging behavior. *Journal of School Psychology*, *49*, 529-554.

Myers, D., Simonsen, B., & Sugai, G. (2011). Increasing teachers' use of praise with a response to intervention approach. *Education & Treatment of Children*, *34*, 35-59.

Nelson, J. R., & Roberts, M. L. (2000). Ongoing reciprocal teacher-student interactions involving disruptive behaviors in general education classrooms. *Journal of Emotional and Behavioral Disorders*, *8*, 27-37.

Oakes, W. P., Lane, K. L., Jenkins, A., & Booker, B. B. (2013). Three-tiered models of prevention: Teacher efficacy and burnout. *Education & Treatment of Children*, *36*, 95-126.

Oliver, R. M., Wehby, J. H., & Reschly, D. J. (2011). Teacher classroom management practices: Effects on disruptive or aggressive student behavior. *Campbell Systematic Reviews*, *4*, 1-55.

Reinke, W., & Herman, K. (2002). Creating school environments that deter antisocial behaviors in youth. *Psychology in the Schools*, *39*, 549-559.

Reinke, W., Lewis-Palmer, T., & Merrell, K. (2008). The classroom check-up: A classwide teacher consultation model for increasing praise and decreasing disruptive behavior. *School Psychology Review*, *37*, 315-332.

Ross, S. W., & Horner, N. (2007). Teacher outcomes of School-Wide Positive Behavior Support: A case study. *Teaching Exceptional Children Plus*, *6*, Article 6.

Ross, S. W., Romer, N., & Horner, R. H. (2012). Teacher well-being and the implementation of school-wide positive behavior interventions and supports. *Journal of Positive Behavior Interventions*, *14*, 118-128.

Satterthwaite, F. E. (1947). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, *2*, 110-114.

Sørlie, M.-A., & Ogden, T. (2014). Reducing threats to validity by design in a nonrandomized experiment of a school-wide prevention model. *International Journal of School & Educational Psychology*, *4*, 235-246.

Sørlie, M.-A., & Ogden, T. (2015). School-Wide Positive Behavior Support–Norway: Impacts on problem behavior and classroom climate. *International Journal of School & Educational Psychology*, *3*, 202-217. doi:10.1080/21683603.2015.1060912

Sørlie, M.-A., & Torsheim, T. (2011). Multilevel analysis of the relationship between teacher collective efficacy and problem behavior in school. *School Effectiveness and School Improvement*, *2*, 175-191.

Sprague, J. R., & Walker, H. M. (2005). *Safe and healthy schools: Practical prevention strategies*. New York, NY: Guilford Press.

Statistics Norway. (2009). Education statistics. Elementary and lower secondary schools. Retrieved from http://www.ssb.no/english/subjects/04/02/20/utgrs_en/tab- 2009-04-29-05-en.html

Sugai, G., Horner, R. H., & Todd, A. W. (2009). *Effective behavior support (EBS) Self-Assessment Survey* (Version 3). Eugene: University of Oregon, Educational and Community Supports. Retrieved from https://www.pbisapps.org/Applications/Pages/PBIS-Assessment-Surveys.aspx#sas

Sutherland, K. S., & Oswald, D. P. (2005). The relationship between teacher and student behavior in classrooms for students with emotional and behavioral disorders: Transactional processes. *Journal of Child and Family Studies*, *14*, 1-14.

Todd, A. W., Campbell, A. L., Meyer, G. G., & Horner, R. H. (2008). The effects of a targeted intervention to reduce problem behaviors: Elementary school implementation of check in-check out. *Journal of Positive Behavior Interventions*, *10*, 46-55.

Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W. K. (2008). Teacher efficacy: Its meaning and measure. *Review of Educational Research*, *68*, 202-248.

Author Biographies

Mari-Anne Sørlie has background in special education and is a senior researcher at the Norwegian Center for Child Behavioral Development. Her primary research interests are evaluation of school-based interventions, and the development of behavior

problems and social competence in children and youths. She is the author of several research articles and has contributed to pedagogical books and reports, written both in Norwegian and English.

Terje Ogden (PhD) holds the position as research director at the Norwegian Center for Child Behavioral Development and is also professor at the Institute of Psychology, University of Oslo. Research interests include clinical trials and implementation of empirically supported interventions (ESI) targeting antisocial children and youth, and the longitudinal development of social competence and externalizing problem behavior in early childhood. Ogden has published several articles on interventions like *Multisystemic Therapy, Parent Management Training – the Oregon model*, and *Early Interventions for Children at Risk*. See www.ogden.no

Asgeir Røyrhus Olseth has a master's degree in educational research methodology and works as a research coordinator at the Norwegian Center for Child Behavioral Development. His main research interests are psychometrics and the effectiveness of social and psychological interventions for children and youths.

IV

World Journal of Educational Research
ISSN 2375-9771 (Print) ISSN 2333-5998 (Online)
Vol. 2, No. 2, 2015
www.scholink.org/ojs/index.php/wjer

Preventing Problem Behavior in School through School-Wide Staff Empowerment: Intervention Outcomes

Mari-Anne Sørlie[1*], Terje Ogden[1] & Asgeir Røyrhus Olseth[1]

[1] Norwegian Center for Child Behavioral Development (NCCBD), Oslo, Norway

[*] Mari-Anne Sørlie, E-mail: m.a.sorlie@atferdssenteret.no

Abstract

The aim of this study was to examine the effects of the universal "Preventing Problem Behavior in School" (PPBS) intervention on both establishing high-quality learning environments and increasing the use of positive teaching strategies to prevent student problem behavior. PPBS was developed and piloted in Norway as an abbreviated version of the School-Wide Positive Behavior Support Model (SWPBS) and includes a four-day in-service training program for a school's entire staff. Seventeen primary schools (Grades 1-7) implementing PPBS and 20 control schools engaging in "practice as usual" were compared using a three-wave measurement design. Multilevel analyses based on staff ratings indicated significant positive main effects of PPBS in the moderate range on the level of school behavior problems, positive behavior management, and perceived staff efficacy. Moreover, school size, implementation quality, proportion of unqualified staff members, and program training dosage moderated the intervention outcomes. Student ratings did not, however, support the staff ratings. The results are discussed in relation to the outcomes of the full-scale SWPBS model, meta-analyses of school-wide interventions, and measurement issues. Study limitations, strengths, practical implications, and future directions are highlighted.

Keywords

problem behavior, school-wide intervention, multi-level analysis, non-randomized experiment

1. Introduction

Aggressive and disruptive behaviors such as bullying, fighting, and arguing, along with a cluster of activities that disturb learning such as unrest, talking out loud and "mental absence" during lessons, have been rated by Norwegian school leaders as among the greatest challenges in today's schools (e.g., Ogden, Sørlie, Arnesen, & Meek-Hansen, 2012). Problematic student behavior reduces the quality of the learning environment, inhibits positive social relationships with classmates and teachers, and reduces the time teachers spend teaching, which in turn negatively affects students' academic performance, school attachment, and social well-being (e.g., Houts, Caspi, Pianta, Arseneault, & Moffitt, 2010; Simonsen et al., 2012). Moreover, there is ample empirical evidence that children showing high levels of externalizing behavior problems in primary school are at significant risk of

developing a persistent antisocial and maladaptive path, particularly if they have additional academic problems (e.g., Mytton, DiGuiseppi, Gough, & Logan, 2007; Duncan & Magnuson, 2011).

Research also indicates that teachers' beliefs in both their own efficacy (self-efficacy) and their collective efficacy are positively related to student learning and teaching performance while negatively and bi-directionally related to problematic student behavior (e.g., Bandura, 1997; Goddard, 2002; Klassen & Tze, 2014; Skaalvik & Skaalvik, 2007; Sørlie & Torsheim, 2011). It has been observed that teachers who express lower self-efficacy are less tolerant of problem behavior, more likely to use punitive and reactive discipline, and more likely to exclude challenging students (Jordan & Stanovich, 2003). The ability of teachers to organize classrooms and to manage student behavior is related not only to student outcomes but also to high perceived teacher efficacy and perseverance (Gibbs & Powell, 2011). Therefore, the ability of school-based interventions to increase teachers' efficacy perceptions should be an important indicator of effective practice in the prevention and reduction of student problem behavior.

For more than 40 years, Norway's primary school-political objective has been to fully include all students and to reduce the use of segregated special education. But between 2007 and 2011, training in segregated settings increased by 30% and students with externalizing behavior problems have proven particularly difficult to include (Ogden, 2014). Although the level of problem behavior now appears to be declining slightly (e.g., Sørlie & Ogden, 2014; Wendelborg, 2011), more effective, systematic, inclusive, and preventive school interventions are necessary (e.g., NOU, 2015). Numerous intervention models and treatment programs have been validated by research, few of them have, however, been widely disseminated to organizations and users eligible for those interventions (Glasgow, 2009). In search of new initiatives to successfully mainstream special needs students, in general, and students with emotional and behavioral difficulties, in particular, the Norwegian Center for Child Behavioral Development (NCCBD) found that the School-Wide Positive Behavior Support (SWPBS) model matched the principles of inclusive education and a promising approach to the prevention of student misconduct.

Concerns about the underutilization of empirically supported interventions and their modest ability to reach potential target groups have been expressed by several scholars (e.g., Fixsen, Naoom, Blasé, Friedman, & Wallace, 2005; Palinkas & Soydan, 2012). In a recent article, Rotheram-Borus, Swendeman and Chorpita (2012) claimed that many of the existing intervention programs may "overserve" the majority of users and that users' essential needs could be met by alternatives that are both less time-consuming and less expensive. The large-scale implementation of evidence-based or empirically supported interventions in schools—such as the SWPBS model—is primarily focused on preventing student problem behavior, and this model may be both simplified and more broadly adapted in less time and at a lower cost. Since 2002, the SWPBS model (called PALS in Norway) has been implemented in 215 of 2,886 Norwegian elementary and secondary schools (7.5%). Thus, although there is a greater potential for full-scale implementation of the SWPBS/N-PALS model, some schools

can most likely manage with a model that is less comprehensive but nevertheless beneficial to both staff and students. With reference to Rotheram-Borus et al. (2012), Jones and Bouffard (2012, p. 13) have argued that such innovations have "rarely replaced the more intensive options but instead reached a wider population, particularly those who would not traditionally have been served". To investigate whether an abbreviated version of the N-PALS model could be sufficient for some schools, the universal "*Preventing Problem Behavior in School*" (PPBS) intervention was developed at NCCBD (Arnesen & Meek-Hansen, 2011) and tested in a non-randomized experiment. In this paper, we present results of the effects of the PPBS intervention for both staff and students.

1.1 School-Wide Interventions

Unlike individually oriented approaches to misconduct and academic problems, which restrict their focus to small groups or individual students at high to moderate risk of conduct problems, school-wide interventions such as the SWPBS/N-PALS model typically have an universal focus and prioritize on changing the social and organizational aspects of entire schools to promote positive student outcomes and prevent problem behavior. These interventions also differ from interventions that focus solely on classrooms, for example, classroom management, in which the focus is restricted to what takes place in class. Several meta-analyses of school-based interventions emphasize the relevance and preventive effects of universal school-wide approaches (e.g., Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011; Weare & Nind, 2011), of which SWPBS has the greatest catchment area.

More specifically, the SWPBS/N-PALS model can be described as a structured and comprehensive (but flexible) system-level approach not only to prevent and reduce student problem behavior but also to promote a positive and inclusive school environment that can facilitate teaching, optimal learning, and psycho-social functioning for all students (Arnesen & Meek-Hansen, 2011). It focuses on systematic, supportive, collective (school-wide), proactive, and inclusive practices (Ogden et al., 2012; Sugai & Horner, 2009). The *full-scale* intervention model has a multilevel structure and takes 3-5 years to fully implement. Based on assessments of student risk levels and the school's needs, the staff successively implements a continuum of evidence-based interventions at the universal level (targeting all students and staff), the selected level, and the indicated level. Interventions at the selected level are designed for the 5-10% of students at moderate risk of conduct problems and school failure who barely respond to interventions at the universal level (Muscott, Mann, & LeBrun, 2008). The indicated level targets the 1-5% of high-risk students who require more intensive and comprehensive interventions than the interventions offered at the first two levels.

The core components of the full-scale SWPBS/N-PALS model are as follows: 1) school-wide positive behavior support strategies, including the teaching of 3-5 positively formulated school rules and systematic praise and encouragement of positive behavior; 2) monitoring of student behavior in all areas of the school using a Web-based assessment and evaluation tool (School-Wide-Information System, SWIS); 3) collectively applied school-wide corrections of problem behavior with mild and immediate consequences (response cost); 4) time-limited small-group instruction or training in

academic or social topics; 5) individually tailored interventions and support plans; 6) classroom management skills for teachers; and 7) parent information and collaboration strategies. The core components, basic training, and implementation features are the same in the USA and Norwegian versions. Except for minor adaptations of the training materials, no changes were made to the original model when SWPBS was transported to the Norwegian context (for more details, see Arnesen, Meek-Hansen, Ogden, & Sørlie, 2014; Ogden et al., 2012).

1.2 Research on SWPBS

Numerous studies of the effects and challenges of the implementation of the SWPBS model have been conducted, primarily in the USA. Overall, the findings after one to two years of implementation indicate several benefits: a) reduced office-discipline referrals, suspensions, tardiness, aggression, and concentration problems (e.g., Bradshaw, Mitchell, & Leaf, 2010; Bradshaw, Pas, Goldweber, Rosenberg, & Leaf, 2012; McIntosh, Bennet, & Price, 2011; Waasdorp, Bradshaw, & Leaf, 2012); b) increased prosocial behavior and emotional regulation (e.g., Bradshaw et al., 2012); and c) improved school climate, as indicated by positive teacher-student relations, student relations and increased school safety (e.g., Backenson, 2012), particularly in schools with high implementation quality scores (e.g., Dix, Slee, Lawson, & Keeves, 2012).

With respect to the outcomes of the full-scale SWPBS/N-PALS model in Norway, research findings indicate moderate to substantial positive effects on a) the occurrence of severe and moderate problem behaviors within and outside classrooms after two years (Sørlie & Ogden, 2007) and three years of implementation, respectively (Sørlie & Ogden, 2015); b) social skills and academic performance in students with immigrant backgrounds (Ogden, Sørlie, & Hagen, 2007); c) classroom climate； and d) inclusion (Sørlie & Ogden, 2015). Additionally, recent analyses have revealed positive effects on school employees'； e) perceived self-efficacy； f) collective efficacy, and g) behavior-management practices (Sørlie, Ogden, & Olseth, in review).

Two user surveys support the results of the effectiveness studies. In one study, 90% of all principals with one to nine years of experience with N-PALS expressed great satisfaction and reported positive outcomes (Sørlie, Ogden, Arnesen, Olseth, & Meek-Hansen, 2014) although some principals reported that N-PALS was a comprehensive and demanding intervention model to implement. The second survey showed that compared with schools implementing other evidence-based programs (e.g., the Olweus bullying prevention program; Olweus & Limber, 2010), the N-PALS schools expressed significantly higher user satisfaction (Eriksen, Hegna, Bakke, & Lyng, 2014).

Valid evidence of positive and multiform effects of the SWPBS model is demonstrated across countries and continents. On the other hand, limited diffusion, and feedback indicating that the model is too comprehensive or demanding for some schools inspired the development and testing of a less-exigent version (PPBS).

1.3 The PPBS Intervention Versus the N-PALS Model

The PPBS intervention was based on the same principles as the full-scale N-PALS model. But, unlike the three-level N-PALS model, the abbreviated PPBS intervention focused only on the universal level including school-wide positive behavior support strategies with 3-5 positively formulated expectations, systematic praise and encouragement of positive behavior, and collectively applied school-wide corrections of problem behavior. More precisely, the key features were 1) a school-wide approach and differentiated evidence-based practices, 2) systematic positive reinforcement of expected pro-social behavior, 3) corrections (mild consequences) following problem behavior, 4) good directions, and 5) establishing a functional support system. The PPBS included a 30-hour in-service training program for school staffs that lasted four full days (two in the autumn and two during the spring semester) and was locally organized with 1-7 schools per site. The entire school staff was included in the program training (i.e., the principal, teachers, assistants, special education teachers, social workers, after-school personnel, and representatives of the school's psychological service), which was provided free of charge. The participants were provided with an intervention manual, and all of the training materials could be downloaded from the Internet. The standardized training sessions were led by the program developers and were composed of a combination of lectures, demonstrations, training, coaching and "homework".

By comparison, in the full-scale version (N-PALS), an internal school-implementation team (5-7 persons) is locally trained by a certified coach for 40 hours over two years, followed by two one-day booster sessions per year. This team is responsible for the developmental work at their school and for informing and training their colleagues in key model features. In addition, unlike in the full-scale version, the PPBS schools did not have access to the SWIS component or received external supervision, training in interventions relevant for students at risk, or any technical support. Different from N-PALS schools, the PPBS schools did not use systematic data on discipline referrals or functional behavioral assessments to determine appropriate interventions. Nor did they have access to either measures of the implementation quality or Web-based information to evaluate their outcomes and progress.

1.4 Research Questions

The research questions of the current study were: a) To what extent does the PPBS intervention affect the prevalence of student problem behavior and the classroom climate (as rated by staff and students) over time?; b) To what extent does the PPBS intervention affect the school staff's disciplining practices (i.e., use of positive behavior supports and behavioral corrections as rated by staff and students) and their perceived self-efficacy and collective efficacy?; c) To what degree are the intervention outcomes moderated by school size, proportion of unqualified staff, program training dosage, and implementation quality?

2. Method

Intervention outcomes were evaluated using questionnaire data from principals, school staffs, and students in a three-wave, nonrandomized experiment in which 17 intervention schools and 20 control schools continuing their "regular practice" participated. An open-cohort design was applied, enabling the enrollment of new staff and students at each measurement point. Participants who changed (or left) schools or began junior high school during the study were not followed.

2.1 Participants

2.1.1 School and Student Characteristics

The mean school size was 338 students (varying from 89 to 780). At baseline (T1), there were 11,367 students in 1st to 7th grades (PPBS=5,606, 49% boys; control=5,761, 52% boys), of whom 12% had minority backgrounds. Approximately 5% received special education services (national level 5.5%), 4% had been referred to educational-psychological services, and fewer than 2% had been referred to child-welfare or mental-health services. Few students had been expelled from school (0.3%), transferred to another school or class (0.2%), or reported to the police (0.2%) because of their challenging behavior.

2.1.2 Staff Characteristics

At baseline, there were 675 employees in the intervention group and 658 in the control group. Seventy-three percent were employed as teachers. The rest of the staff members were assistants, after-school personnel and school administrators. Eighty percent were female, and most were middle-aged (73% older than 35 years). The teachers were experienced (only 19% had worked at the school for fewer than five years) and well educated. Four percent had no formal training, and 11% had special education training. Seventy percent worked full time.

2.1.3 Baseline Comparisons

Unconditional analysis at baseline showed that the control group reported significantly higher perceived collective efficacy than the PPBS group, $F(1,957)=23.81$, $p<0.05$, whereas the level of problem behavior occurring outside the classroom, $F(1,899)=10.87$, $p<0.05$, school size, $F(1,972)=150.49$, $p<0.05$, and number of unqualified staff, $F(1,972)=44.53$, $p<0.05$, were higher in the PPBS group. The baseline comparisons indicated that initially, the situation in the intervention schools was more challenging than in the control schools. Nevertheless, the group differences were in the small-to-moderate range, and by controlling for initial variation on observed variables, the outcome analyses allowed for meaningful group comparisons.

2.2 Procedures

The schools were recruited from two strategically selected pools of municipalities. Schools actively implementing other school- or community-wide programs, such as the Olweus bullying-prevention program (Olweus & Limber, 2010) or the TIBR program (Kjøbli & Sørlie, 2008), were excluded as potential participants to avoid program contamination. Recruitment into the intervention group was conducted by an open invitation to all elementary schools in actual northern and southern

municipalities. The eligible sample was composed of 48 schools, of which 17 agreed to participate. A random sample of 44 of the 126 eligible schools located in the actual western municipalities was invited as controls; 20 agreed to participate. Comparing the refusing and participating schools by reference to nationally standardized achievement scores (math, reading and English performance, 5th grade) and data from 7th graders related to school learning environments (annual Student Survey) revealed no group differences.

Data for the current study were collected in three waves: Time-point one (T1, baseline, staff only) was at the end of the school year prior to the implementation of the PPBS. Time-point two (T2) was six months later, at the beginning of a new school year and close to the initiation of the intervention. Time-point three (T3, post-test) was at the end of the school year and four months after the program training; we expected the schools to require some time to implement the intervention. Questionnaires were completed during ordinary class or working time and were available both on the Internet and on paper. Written instructions were presented to standardize the assessment procedures, and consent from both staff and parents were obtained in advance. Only staff who were in daily and direct contact with a group of students were asked to participate, resulting in an actual sample of N=1,266. The school leaders participated at T1 only by completing a questionnaire about their schools' structural and organizational characteristics.

2.3 Measures

2.3.1 Problem Behavior

Student problem behavior was measured by staff ratings using two British scales (Grey & Sime, 1989) "Problem Behavior in the School Environment" (15 items) and "Problem Behavior in the Classroom" (20 items). The staff reported how many times they had observed negative incidents inside and outside the classroom during the week prior to assessment (i.e., a random week). Item examples include "Running in corridors" and "Physical attacks on students". A 5-point Likert scale was applied, with scores ranging from 1 (*not observed*) to 5 (*observed several times per day*). The scales have shown satisfactory psychometric properties in prior Norwegian studies (e.g., Sørlie & Ogden, 2007; Kjøbli & Sørlie, 2008). The Cronbach's alphas in the present study ranged from α=.81 to .88 across assessment points. Factor analysis revealed two underlying sub-factors for each measure: "*Severe problem behavior*" (6 and 12 items) and "*Moderate problem behavior*" (9 and 8 items). The sub-factors had acceptable internal reliability (α=.72 to .89), except for "Severe problem behavior in common areas" (α=.43 to .48).

2.3.2 Classroom Climate

The "Classroom Environment Scale" (CES; Moos & Trickett, 1974) was used by staff members to assess the quality of the general learning conditions in class. The CES is a 14-item scale (α ranged from .82 to .83) that includes statements such as "The students in this class help each other" and "Usually, the students finish ordered working tasks". An equivalent 22-item student scale (Sørlie & Nordahl, 1998) assessed the students' perceptions of the psychosocial learning conditions in their

classrooms (α ranged from .86 to .88). A 4-point rating scale (1=*does not fit*, 4=*fits completely*) was used. Satisfactory psychometric properties have been shown in prior Norwegian studies (e.g., Sørlie & Ogden, 2007, 2015; Sørlie & Nordahl, 1998).

2.3.3 Collective Efficacy and Self-Efficacy in Schools

Perceived staff collective efficacy was measured with a 12-item scale developed by Goddard (2002). Cronbach's alpha ranged from .95 to .96 across assessment points in the current study. Items were rated on a 5-point scale ranging from 1 (*never*) to 5 (*very often*) and included statements such as "Teachers here are confident that they will be able to motivate their students" and "Teachers in this school are able to get through to difficult students". Self-efficacy was measured with a 30-item scale (Sørlie & Ogden, 2015; α ranged from .95 to .96). The employees rated how competent they felt in managing and preventing problem behavior and promoting academic skills. The rating scale ranged from 1 (*highly incapable*) to 7 (*highly capable*) on items such as "To stop aggressive student behavior", and "To be present in the classroom when the lesson begins". Factor analysis unveiled the expected two-factor structure: "Ability to prevent problem behavior" (9 items, α from .86 to .88) and "Ability to bring about student learning" (21 items, α from .94 to .95).

2.3.4 Behavior Management

Strategies to promote prosocial behavior and manage problem behavior were assessed using a 17-item staff scale and an equivalent 23-item student scale (Sørlie & Ogden, 2015). For both scales, factor analysis revealed a two-factor structure: "Positive behavior support" (staff: 9 items, α from .74 to .76; students: 16 items, α from .88 to .91) and "Behavioral correction" (staff: 8 items, α from .61 to .63; students: 5 items, α from .53 to .64). The staff reported how many times they had used strategies such as "Praised and encouraged expected positive student behavior" and "Deliberately ignored undesirable behavior" during the previous month on a scale ranging from 0 (*0 times*) to 6 (*more than 20 times*). The students reported how often in the previous month they had experienced teacher practices such as "The teacher praises me when I do as she/he says during lessons" and "The teacher says what will happen if we violate the rules" (1=*never*, 4=*usually*).

2.3.5 Implementation Quality

Implementation quality was measured in the PPBS schools at post-test with a 15-item short version (α=.74) of "The Effective Behavior Support Self-assessment Survey" (EBS-SAS), developed by Sugai, Horner, and Todd (2009). The EBS-SAS has shown satisfactory psychometric qualities in several prior evaluation studies (e.g., Bradshaw et al., 2012; Sørlie & Ogden, 2015). The staff was asked how various statements (e.g., "Expected student behavior is consequently encouraged and positively acknowledged", "A few (3-5) school rules are positively and clearly defined", "Problem behavior is consistently addressed with mild and predictable negative consequences", and "The staff members who have inspection actively supervise the students in all arenas outside the classroom context") corresponded with the situation at their schools on a 3-point scale (1=*fits completely*, 3=*does not fit*).

2.3.6 Dosage

PPBS training dosage refers to the intervention school's mean training attendance score. The score was calculated on each staff member's participation across the four training days and was aggregated at the school level. The correlation between implementation quality and dosage was significant, but in the moderate range with r=.24 ($p<.01$).

2.4 Analytic Strategy

A longitudinal, multilevel model was used to examine differences in change over time between the control and intervention groups. The Mixed Linear procedure in IBM SPSS statistics (version 20) was used to accommodate the hierarchical data structure (level 1, T1, T2, T3; level 2, staff or students; level 3, schools). The control group was established as the reference group and T1 as the reference time point, whereas T2 and T3 were included in the analyses. Because of the nonrandomized design, selection bias may represent a threat to internal validity; i.e., it can be questioned whether observed changes in outcome variables over time are caused by the intervention (Shadish, Cook, & Campbell, 2002). To reduce initial school variance, school size, portion of unqualified staff, proportion of special education students, and proportion of students with foreign backgrounds were included as covariates at the school level. To reduce multicollinearity, all covariates were centered (Graham, 2003), and non-normally distributed variables were log-transformed. An unstructured residual covariance structure was chosen for the level 1 residuals, whereas at the school level, random intercepts were estimated using a scaled-identity covariance structure. Missing data were estimated using the direct-likelihood method assuming a missing at-random mechanism (MAR, see below) (Beunckens, Molenberghs, & Kenward, 2005). Main effects were investigated by adding a Time x Group interaction to the models. Moderating effects were investigated by adding the three-way interaction terms Time x Group x School size and Time x Group x Portion of unqualified staff or the two-way interactions Time x Dosage and Time x Implementation quality. In all of the analyses, the lower-order terms were included to ensure balanced regressions equations. One PPBS school had an exceptionally low mean training dosage (32.3%), and two control schools resigned from the study prior to T2 and T3 because of task overload. However, all of the schools were included in the analyses in accordance with the intention-to-treat (ITT) principle. Effect sizes (Cohen's d) were calculated according to Feingold's (2013) recommendations.

3. Result

3.1 Attrition

At the time of the first assessment, 77% of staff members in the PPBS and control schools were participating. Overall, 1,308 staff members participated at one or more time points. The pre-post attrition was attributable to a) two schools withdrawing from the study (n=89), b) change of workplace (n=246), c) short-term leaves of absence (n=84), d) no longer work with groups of students (n=71), and e) unwillingness to participate (n=36). Comparing the participants at baseline with those missing at post-test revealed no differences in age, gender or class level. Fewer in the missing group were trained

as teachers compared with those who participated, $F(1,997)=21.97$, $p<0.01$, and more participants in the missing group were less experienced, $F(1,993)=-2.58$, $p<0.01$. Overall, there were 6,172 students (PPBS=3,087, control=3,085) in grades 4-7; the parents of 14% of those students did not allow them to participate; 9.9% never returned the consent form.

Evenly distributed across grades, 4,687 students contributed to the study at pre-test and 4,630 contributed at post-test. The response rates were high (95.5% and 91.5%). By post-test, 221 students were missing because two control schools withdrew from the study, 102 were new students, 117 had changed schools, and consent was withdrawn for four students. Except for somewhat lower pre-test ratings of the teacher's use of positive behavior support among those missing at post-test (M=54.83 compared with 55.91, t=-2.27, p=.023), no significant differences between the missing and participating students were identified.

In conclusion, attrition in this study was modest and reflects its open-cohort design, two schools dropping out, normal fluctuations in the staff and student populations, and few differences between participants and non-participants were identified. Accordingly, missing data were estimated using the direct-likelihood method (Beunckens et al., 2005).

3.2 Indications of Main Intervention Effects

To simplify the results of multilevel analysis, which are often difficult to read, the estimates of the fixed effects in Table 1 are shown as group differences in change across measure points. Effect sizes (Cohen's *d*) with 95% confidence intervals are specified. A positive *d* signifies a positive (desired) intervention effect. The intervention effects were in the small to modest range, with *d* from .10 to .41.

Table 1. Main Effects of PPBS. Change Across Time within and between the PPBS and Control Group. Fixed Effects Estimates and Effect Sizes (Cohen's d) with Confidence Intervals

Change within groups		
	PPBS	Control
Variable	Pre -post	Pre -post
Problem behavior on common areas	-2.95	-0.42
- Moderate problem behavior	-2.43	-0.20
- Serious problem behavior	-0.50	-0.21
Problem behavior in classroom	-2.98	-1.81
- Moderate problem behavior	-2.15	-1.40
- Serious problem behavior	-0.81	-0.35
Classroom climate (S)	0.70	0.30
Classroom climate (St)	-1.78	-2.12
Collective efficacy	1.25	0.46
Self-efficacy	3.06	1.22

Positive behavior support (S)	1.95	-0.95
Positive behavior support (St)	-0.29	-0.11
Behavioral correction (S)	-0.97	-0.66
Behavioral correction (St)	-0.01	-0.06

Difference between groups			
		ES	
Pre -post	Sign.	*d*	*d* 95% CIΔ
-2.50	.000	0.38	0.375- 0.385
-2.23	.000	0.41	0.405 - 0.415
-0.28	.020	0.18	0.175 - 0.185
-1.17	.058	0.13	0.125 - 0.135
-0.75	.121	0.11	0.105 - 0.115
-0.46	.037	0.15	0.145 - 0.155
0.40	.288	0.08	0.066 - 0.074
0.34	.172	0.04	0.039 – 0.040
0.79	.063	0.12	0.124 - 0.116
1.84	.062	0.10	0.104 - 0.096
2.90	.000	0.41	0.414 - 0.406
-0.18	.570	0.02	0.019 – 0.021
-0.31	.163	0.09	0.094 - 0.086
0.05	.681	0.02	0.019 – 0.021

Note. Estimates (enhanced values) based on Satterwaite's (1947) approximate degrees of freedom. All covariates and the Group x Time interaction are accounted for in the estimates of change. S=staff, St=student.

Indications of a main effect of the PPBS intervention (i.e., effect for all schools) were observed in staff-reported problem behavior in common school areas (T1-T3 diff=-2.50, p=.000). There was a decrease across time in both groups (PPBS T1-T3 change=-2.95, control T1-T3 change=-0.42); however, a significantly greater reduction occurred in the intervention group than in the control group. This effect was true both for severe (T1-T3 diff=-0.28, p=.020) and moderate problem behaviors (T1-T3 diff=-2.23, p=.000). The prevalence of problem behaviors observed in the classrooms was also substantially reduced during the intervention period (PPBS T1-T3 change=-2.98, control T1-T3 change=-1.81), and indications of a marginally significant PPBS effect were found (T1-T3 diff=-1.17, p=.058). This reduction across time was primarily related to reduced occurrence of severe problem behaviors (T1-T3 diff=-0.46, p=.037). No main effect on the classroom climate was found, either as

rated by staff or as rated by students.

That notwithstanding, the multilevel analyses based on staff reports indicated a significant main effect on disciplinary practices in the schools. The staff at the PPBS schools increased the number of positive behavior supports from baseline to post-test more than did their colleagues in the control schools (T1-T3 diff=2.90, *p*=.000). Although there was a general decrease in the use of behavioral corrections, no significant group difference was found. Likewise, student ratings of the teachers' behavior management showed no significant group differences. However, the analysis indicated a marginally significant main effect on collective efficacy (T1-T3 diff=0.79, *p*=.063). A more positive change was registered in the intervention schools (T1-T3 change=1.25) than in the control schools (T1-T3 change=0.46. During the study period, there was also a general increase in perceived self-efficacy (PPBS T1-T3 change=3.06, control T1-T3 change=1.22); however, the analysis only indicated a marginally significant more positive trend in the PPBS group than in the control group (T1-T3 diff=1.84, *p*=.062).

3.3 Indications of Differential Effects

Parameter estimates from the additional moderation analyses are shown in Table 2. School size and program-training dosage significantly moderated the intervention effects on three outcome variables, and the portion of unqualified staff and implementation quality moderated the effects on two outcome variables. No moderation effects were found on student-rated variables.

To investigate the differential effects of PPBS more closely, we divided the schools into three groups according to Norwegian school size (small=n<200 students, medium=201-350 students, large=351-780 students) and two groups according to proportion of unqualified staff, program training dosage, and implementation scores (1=high, i.e., above the mean, 2=low, i.e., below the mean). In Norway 29.3% of the schools have 300 or more students, 40% have between 100 and 299 students, and 30.7% have less than 100 students.

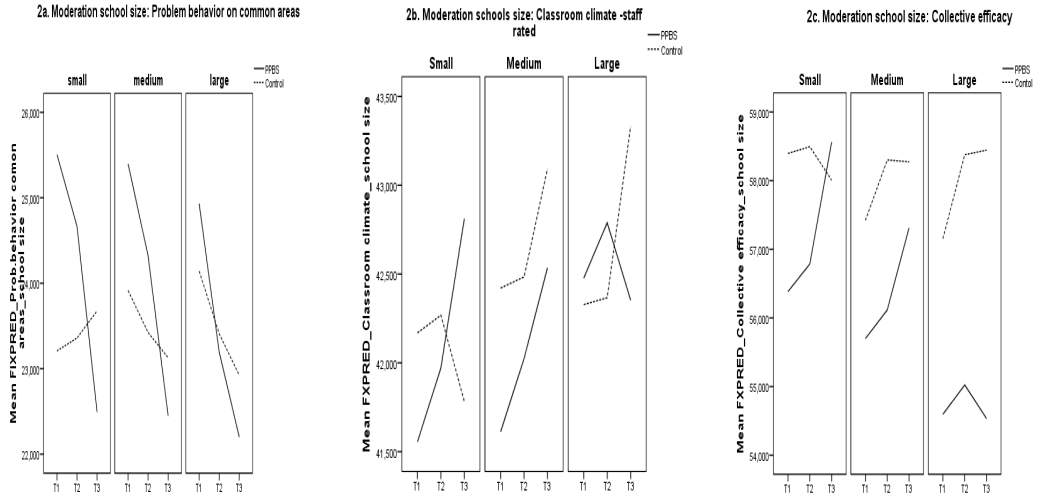
Table 2. Significant Moderating Effects on PPBS Outcomes of School Size, Portion of Unqualified Staff, Program Training Dosage, and Implementation Quality

Outcome variable	Moderation variable	Estimate (SE)	df	t	Sign.
Problem behavior on common areas	School size	.012 (.005)	884.17	2.40	.017
	Dosage	-.037 (.012)	478.80	-1.91	.057
Classroom climate (S)	School size	-.013 (.004)	915.93	-3.41	.001
	Unqualified staff	.126 (.052)	903.12	2.44	.015
	Implementation quality	.182 (.063)	398.05	2.91	.004
Collective efficacy	School size	-.018 (.005)	844.49	-4.01	.000

	Dosage	.074 (.017)	453.75	4.41	.000
	Implementation quality	.174 (.071)	388.05	2.46	.014
Positive behavior support (S)	Unqualified staff	.171 (.071)	856.01	2.43	.016
	Dosage	.052 (.022)	466.88	2.39	.017

Note. Estimates (enhanced values) based on Satterwaite's (1947) approximate degrees of freedom. The covariates time (T1,T2,T3), total number of students, portion of students with foreign background, portion of special education students, and portion on unqualified staff, program dosage, implementation quality and relevant two and three 3-way interactions were accounted for.

The analyses indicated that when compared with control schools of the same size, the small-to-medium PPBS schools benefited more than larger schools in terms of reduced problem behavior in common school areas, better classroom climate, and higher perceived collective efficacy (Figure 1(2a-2j)).



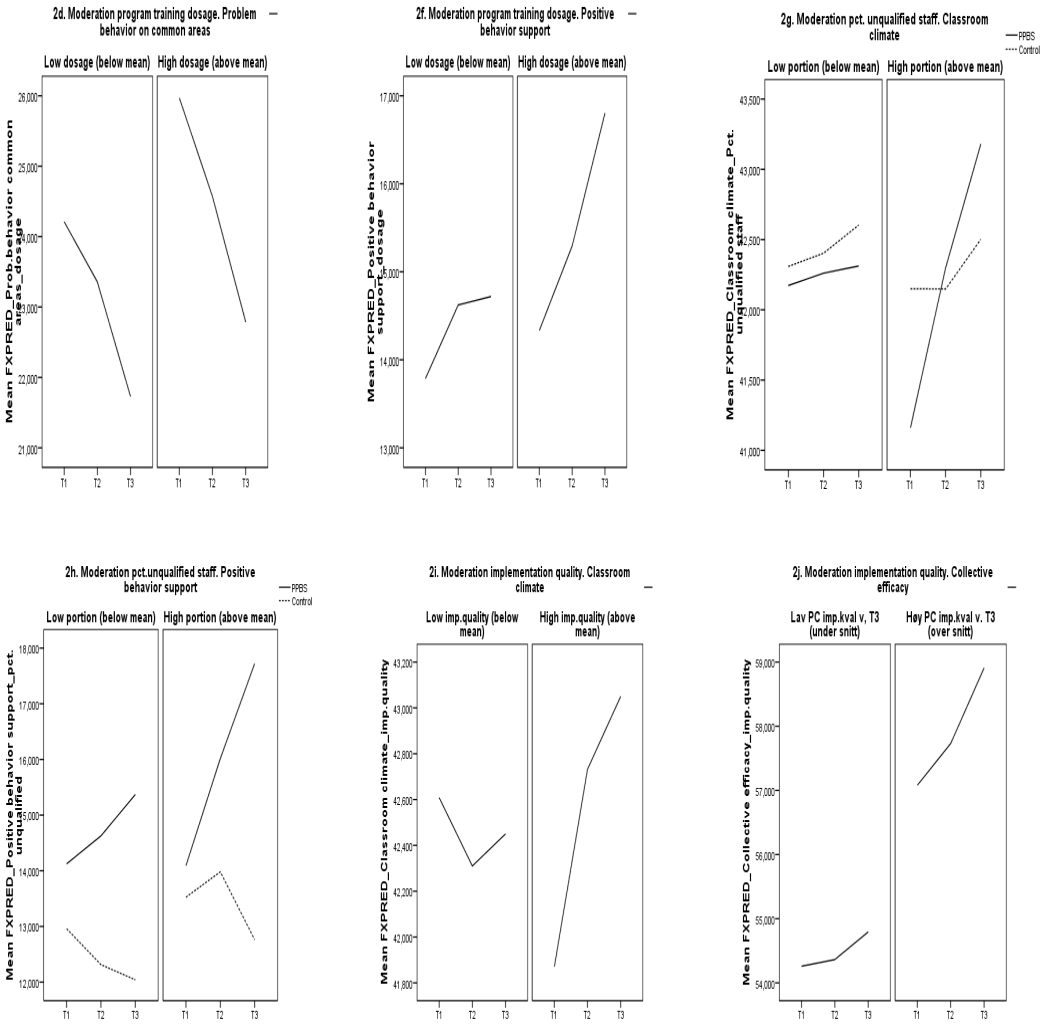


Figure 1. (2a-2j). Moderating Effects of Schools Size, Portion of Unqualified Staff, Program Training Dosage, and Implementation Quality

Moreover, PPBS schools with a high mean training dosage demonstrated a significantly greater reduction in student problem behavior than schools with a low dosage (Figure 1d). Greater increases in collective efficacy were also found in the schools with the highest program dosage (Figures 1e-f). Likewise, intervention schools with high implementation quality showed greater improvement in classroom climate and positive behavior supports from PPBS than did schools with low fidelity scores (Figures 1i-j). Compared with control schools that had varying proportions of unqualified staff, the positive changes across time in both the quality of the learning conditions in classrooms and the staff's use of positive discipline were significantly greater in PPBS schools with high proportion of staff members who lacked formal training as teachers than in schools with lower proportions (Figures 1g-h).

4. Discussion

The primary purpose of the current study was to test the immediate effectiveness of the universal school-based *Preventing Problem Behavior in School* (PPBS) intervention. The aim of this school-wide approach is to prevent and reduce student problem behavior by empowering the entire school staff and motivating them to use more positive, efficient, and co-joint behavior management practices. The PPBS is derived from the multilevel SWPBS model (called PALS in Norway) and was developed as an abbreviated alternative for schools that have neither the need nor the capacity to implement the full model. Multilevel analyses of change across three measurement points in which the 17 intervention schools were compared with 20 schools conducting "practice as usual" indicated interesting main and differential effects of PPBS.

4.1 Effects on Student Behavior

As measured four months after program training and compared with the control schools, the PPBS schools reported substantial reductions in negative behavior incidents occurring *outside* the classroom context, such as in stairways, corridors, bathrooms, dining areas, gymnasiums, and playgrounds (d=.38). Intervention impacts were evident both for less severe problem behaviors—such as unrest while waiting, rude comments to teachers, and lack of care for others (d=.41)—and for more severe problem behaviors—such as theft, vandalism, and physical attacks (d=.18). Moderation analyses indicated greater decreases in student misconduct in small and medium schools than in large schools. It should be mentioned that due to the decentralized structure of the Norwegian school system, a "large" primary school is a school with 300 or more students. The moderation analyses also indicated greater decreases in student misconduct in schools with a high than low program training dosage. A marginally significant main effect was also found for problem behaviors occurring *within* the classroom context (d=.13), and this effect was closely related to decreases in severe problem behaviors (d=.15). None of the context and process variables tested moderated the classroom-related changes.

4.2 Effects on Classroom Climate

No main effect of the PPBS was observed for the quality of the classroom climate as rated by staff or students. However, although not all intervention schools experienced more positive and less conflictual social relations in class following the implementation of the PPBS, the moderation analyses indicated that *some* schools did: schools that implemented the intervention with high fidelity, schools with a larger number of unqualified staff members, and small to medium-sized schools.

4.3 Effects on Staff Behavior

Significant improvements in employees' behavior management following the PBBS were observed, as expressed by staff reports of increased use of behavior-supporting practices such as giving praise, acknowledgement and proactive instructions. The schools with the highest implementation scores and proportion of unqualified staff appeared to benefit most from the intervention. However, according to student reports, no significant group differences in change across time occurred in the staffs' disciplinary practices.

4.4 Effects on Perceived Efficacy

Furthermore, a more positive development in perceived collective efficacy in the intervention group indicated that the staff in the PPBS schools, compared with their colleagues in the control schools, perceived greater confidence in their mutual ability to motivate and support student learning, to reach difficult students, and to prevent and manage problem behavior. The greatest change was observed in the small to moderately sized schools, the schools with the highest fidelity scores, and the schools with the highest program-training dosage. Likewise, a marginally significant main effect was observed in the staff members' self-efficacy beliefs, but there were no differential effects.

4.5 Discrepancies in Staff- and Student-Rated Outcomes

The promising, even modest efficacy indications based on staff ratings were not supported by the students' ratings. Modest between-rater agreement (e.g., between student, teacher and/or parent ratings of child problems) is a familiar phenomenon across studies and countries (Rescorla et al., 2013). Informant discrepancies are also often reported in intervention trials (De Los Reyes, 2011). Whether the discrepancies observed in the present study reflect measurement error, inconsistent intervention effects, inflated staff-based effect estimates, and/or yield useful information for future refinements of the PPBS intervention are unclear. For example, we cannot discount that the more positive staff than student ratings may partially mirror some exaggerated enthusiasm among the staff concerning immediate changes following the implementation of the intervention (i.e., positive response bias). Second, it may be that many students did not experience the institution of more uniform school rules and consequent disciplinary practice as unconditionally positive. Clearer expectations of how to behave in school may feel more restrictive to some students, whereas for other students more frequent positive feedback may feel embarrassing or unfair, particularly early in the implementation and among well-adapted students. Third, we speculate that the students were not particularly sensitive to changes in the *amount* of behavioral feedback they received and that the 0-effects observed in the students' ratings may partially be caused by measurement problems. The reliability of younger students' ratings (9-12 years) can be questioned with respect to how many positive and negative behavioral responses they experienced over a 30-day period, particularly when rated retrospectively, as in this study.

4.6 Contextualization of Results

The significant effect sizes in the current study were in the range of d=.15 to .41, and thereby large enough to be of practical significance. In a prior study, 20 schools implementing the full-scale N-PALS model were compared with the same control group on the same outcome variables as in the present study (Sørlie & Ogden, 2015; Sørlie et al., in review). Four months impacts of the abbreviated PPBS intervention on the level of problem behavior in schools appear relatively equal to three years impacts of N-PALS, whereas the full-scale model appears to "outperform" the abbreviated version concerning effects on the learning conditions in classrooms and on staff outcomes (i.e., on collective efficacy, self-efficacy, and behavior management). The ES-values for PPBS seemed somewhat higher than the ES-values reported for SWPBS in the USA (d from .08 to .17) by Bradshaw et al. (2012). It shold,

however, be noted that the values are not completely compareable in that the American values are based on teacher reports of student behavior as measured at the individual level while the Norwegian ES-values are based on teacher reports of student problem behavior as measured at the class and school level.

Moreover, the outcomes of the PPBS intervention match relatively well with results reported in prior controlled studies with respect to other types of in-service training programs for teachers aimed at improving the social, emotional and behavioral outcomes of primary-school students through improved classroom-management practices (e.g., Whear et al., 2012). The magnitude of the PPBS intervention effects on problem behavior in schools also is consistent with the effects reported in meta-analyses of universal school programs, in which the mean ES-value across a large number of studies is identified as *d*=.20 (Durlak et al., 2011; Wilson & Lipsey, 2007).

4.7 Strengths and Limitations

The primary strengths of the study are the testing of intervention effects within ordinary school settings at no extra cost (free training, materials, location and feeding), a large sample of schools, inclusion of entire school staffs as participants, an experimental double pretest-posttest design, measurement of key constructs using the appropriate units of analysis, multiple informant groups, and the use of multilevel analysis to account for nesting in data. To assess the effects of a school-wide, universal intervention, we found it more relevant to measure school-level than individual- or class-level outcomes.

Because the respondents were not blind to which research condition they had been assigned, the pre-post relations between outcome variables may be inflated because of shared variance from reports of the same informants. Accordingly, we cannot exclude the possibility of biased assessments because of the so-called Hawthorn Effect or because of novelty effects (Shadish et al., 2002). Another limitation relates to the non-randomized design. Although the school groups were more similar than different at baseline, significant group differences on a few observed variables indicated possible selection bias. However, we have intentionally attempted to reduce threats to this study's validity by adding more than one pretest, securing sufficient statistical power to detect small intervention effects, using active and focal controls, and controlling the outcome analyses for initial school variance and potential confounders (Sørlie & Ogden, 2014).

4.8 Conclusions

Considered together, the outcome analyses indicated that PPBS has promising, immediate effects. The analyses suggest greater impacts on problem behavior occurring in common school areas than on problem behavior occurring within the classroom. The analyses also suggest greater impacts on the staff's collective efficacy perceptions than on their individual efficacy beliefs and greater impacts on their use of positive behavior management practices than on their use of correctional strategies to prevent and manage student misconduct. The outcomes may be attributed to the strong emphasis on school-wide and universal components in the PPBS intervention, greater attention to common school premises, and the importance of establishing mutual practices rather than focusing on individual staff

members' adherence to the intervention. It may also be that there is genuinely less potential to change student and employee behavior inside of the classroom than outside of it.

The moderation analyses indicate that to be optimally effective, high attendance by a school's entire staff during training sessions is required and the PPBS intervention must be implemented with high fidelity across all school settings. The moderation analyses also indicate that it is most likely more difficult to induce school-wide changes in large schools than in smaller schools and that more comprehensive implementation supports may be required to "turn large schools around". Moreover, the results indicate that PPBS may positively contribute to increasing school staff's competence and practical skills in the prevention and handling of student problem behavior—particularly among those staff members who do not have formal training as teachers.

Based on the indicative intervention effects, we argue that the PPBS intervention may be appropriate as a basic approach to facilitate the development of school-wide or more-common attitudes and strategies related to the prevention of behavioral misconduct. Earmarking four days to promote effective positive teacher practices should be practicable for most schools. For some schools (e.g., large schools, schools facing stable high levels of problem behavior), although the PPBS may be insufficient, it can function as a "springboard" to more comprehensive and adequate approaches such as the N-PALS full-scale model.

Since the PPBS program is an abbreviated version of N-PALS that was transported from the USA across the Atlantic with only cosmetic adjustments to the original SWPBS model-and also proven effective in Norway as a non-English speaking representative of the Europe countries (Sørlie & Ogden, 2015; Sørlie et al., in review), likely the PPBS can be successively implemented in other contexts and cultures than ours (e.g., USA, Scandinavia). The results must, however be replicated, preferably in randomized controlled studies, before firmer conclusions regarding intervention effects and transportability can be drawn. Additionally, whether the observed intervention effects can be sustained over a longer time span is an open question.

Acknowledgements

The standards of the Regional Committee for Medical and Health Research Ethics (REK South-East) and of the Norwegian Social Science Data Services were followed throughout the conduct of this study. We acknowledge Professor Torbjørn Torsheim for his invaluable support of the analyses. We are also grateful for the intervention adapters, Anne Arnesen and Wilhelm Meek-Hansen, and their genuine interest in the study. None of them was involved in the research.

Funding

This study was supported by the Norwegian Center for Child Behavioral Research and in part by a grant from the Norwegian Directorate for Education and Training.

Conflict of Interest

The authors declare that there is no conflict of interest. None of the authors has been involved in the development, training, or implementation of the N-PALS and PPBS interventions

References

Arnesen, A., & Meek-Hansen, W. (2011). *PALS Håndbok modul 1. Skoleomfattende forebyggende tiltaksnivå*. Oslo: Norsk Senter for studier av problematferd og innovativ praksis (NCCBD).

Arnesen, A., Meek-Hansen, W., Ogden, T., & Sørlie, M-A. (2014). *Positiv læringsstøtte*. *Hele skolen med!* Oslo: Universitetsforlaget.

Backenson, E. M. R. (2012). *Positive Behavior Support and Intervention Programs vs Responsive Classroom Programs: Impact on Perceptions of School Climate*. *PCOM Psychology Dissertations*. Paper Publishing House, p. 238. Retrieved June 14, 2015 from http://digitalcommons.pcom.edu/psychology_dissertations/238

Bandura, A. (1997). *Self-Efficacy: The Exercise of Control*. New York: W. H. Freeman.

Beunckens, C., Molenberghs, G., & Kenward, M. G. (2005). Direct likelihood analysis versus simple forms of imputation for missing data in randomized clinical trials. *Clinical Trials*, *2*(5), 379-386. http://dx.doi.org/10.1191/1740774505cn119oa

Bradshaw, C. P., Mitchell, M. M., & Leaf, P. J. (2010). Examining the effects of school wide positive behavioral interventions and supports on student outcomes. Results from a randomized controlled effectiveness trial in elementary schools. *Journal of Positive Behavior Interventions*, *12*(3), 133-148. http://dx.doi.org/10.1177/1098300709334798

Bradshaw, C. P., Pas, E. T., Goldweber, A., Rosenberg, M. S., & Leaf, P. J. (2012). Integrating school-wide positive behavioral interventions and supports with tier 2 coaching to student support teams: The PBIS plus model. *Advances in School Mental Health Promotion*, *5*, 177-193.

De Los Reyes, A. (2011). Introduction to the special section: More than measurement error: Discovering meaning behind informant discrepancies in clinical assessments of children and adolescents. *Journal of Clinical Child & Adolescent Psychology*, *40*(1), 1-9. http://dx.doi.org/10.1080/15374416.2011.533405

Dix, K. L., Slee, P. T., Lawson, M. J., & Keeves, J. P. (2012). Implementation quality of whole-school mental health promotion and students' academic performance. *Child and Adolescent Mental Health*, *17*(1), 45-51. http://dx.doi.org/10.1111/j.1475-3588.2011.00608.x

Duncan, G. J., & Magnuson, K. A. (2011). The nature and impact of early achievement skills, attention skills, and behavior problems. In G. J., Duncan, & R. J., Murnane (Eds.), *Whither Opportunity: Rising Inequality, Schools, and Children's Life Chances* (pp. 47-69). New York, NY: Russell Sage.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, *82*(1), 405-432.

http://dx.doi.org/10.1111/j.1467-8624.2010.01564.x

Eriksen, I. M., Hegna, K., Bakken, A., & Lyng, S. T. (2014). *Felles focus-en studie av skolemiljøprogrammer i norsk skole*. Rapport 15/2014. Oslo: NOVA.

Feingold, A. (2013). A regression framework for effect size assessments in longitudinal modeling of group differences. *Review of General Psychology*, *17*(1), 111-121. http://dx.doi.org/10.1037/a0030048

Fixsen, D. L., Naoom, S. F., Blasé, K. A., Friedman, R. M., & Wallace, F. (2005). Implementation research: A synthesis of the literature. Retrieved June 14, 2015 from http://www.fpg.unc.edu/~nirn/resources/publications/Monograph/

Gibbs, S., & Powell, B. (2012). Teacher efficacy and pupil behaviour: The structure of teachers' individual and collective beliefs and their relationship with numbers of pupils excluded from school. *British Journal of Educational Psychology*, *82*(4), 564-584. http://dx.doi.org/10.1111/j.2044-8279.2011.02046.x

Glasgow, R. E. (2009). Critical measurement issues in translational research. *Research on Social Work Practice*, *19*(5), 560-568. http://dx.doi.org/10.1177/1049731509335497

Goddard, R. (2002). A theoretical and empirical analysis of the measurement of collective efficacy: The development of a short form. *Educational and Psychological Measurement*, *62*(1), 97-110. http://dx.doi.org/10.1177/0013164402062001007

Graham, M. H. (2003). Confronting multicollinearity in ecological multiple regression. *Ecology*, *84*(11), 2809-2815. http://dx.doi.org/10.1890/02-3114

Grey, J., & Sime, N. (1989). Findings from the national survey of teachers in England and Wales. In *Elton (1989). Discipline in Schools*. Report of the committee of enquiry chaired by Lord Elton. Department of Education and Science and the Welsh office. London: Her Majesty's Stationary Office.

Houts, R. M., Caspi, A., Pianta, R. C., Arseneault, L., & Moffitt, T. E. (2010). The challenging pupil in the classroom: The effect of the child on the teacher. *Psychological Science*, *21*(12), 1802-1810. http://dx.doi.org/10.1177/0956797610388047

Jones, S. M., & Bouffard, S. M. (2012). *Social and Emotional Learning in Schools. From Programs to Strategies Social Policy Report*, *26*, 3-30.

Jordan, A., & Stanovich, P. (2003). Teachers' personal epistemological beliefs about with disabilities as indicators of effective teaching practices. *Journal of Research in Special Educational Needs*, *1*, 1-2.

Kjøbli. J., & Sørlie, M-A. (2008). School outcomes of a community-wide intervention model aimed at preventing problem behavior. *Scandinavian Journal of Psychology*, *49*(4), 365-375. http://dx.doi.org/10.1111/j.1467-9450.2008.00648.x

Klassen, R. M., & Tze, V. M. C. (2014). Teachers' self-efficacy, personality, and teaching effectiveness. A meta-analysis. *Educational Research Review*, *12*, 59-76.

Losel, F., & Bender, D. (2003). Resilience and protective factors. In D. P. Farrington, & J. E. Coid (Eds.), *Prevention of Adult Antisocial Behavior*. Cambridge, UK: Cambridge University Press.

McIntosh, K., Bennet, J. L., & Price, K. (2011). Evaluation of social and academic effects of school-wide positive behavior support in a Canadian school district. *Exceptionality*, *1*, 46-60.

Moos, R. H., & Trickett, E. J. (1974). *The classroom environment scale*. Los Angeles, CA: Consulting Psychology Press.

Muscott, H. S., Mann, E. L., & LeBrun, M. R. (2008). Positive Behavioral Interventions and Supports in New Hampshire: Effects of Large-Scale Implementation of Schoolwide Positive Behavior Support on Student Discipline and Academic Achievement. *Journal of Positive Behavior Interventions*, *10*(3), 190-205. http://dx.doi.org/10.1177/1098300708316258

Mytton, J., DiGuiseppi, C., Gough, D., Taylor, R., & Logan, S. (2007). School-based secondary prevention programmes for preventing violence [Review]. *Evidence-Based Child Health*, *2*(2), 814-891. http://dx.doi.org/10.1002/ebch.127

NOU. (2015). *Å høre til. Virkemidler for et trygt psykososialt skolemiljø. (An official Norwegian report conducted by a committee constituted by the Government)*. Oslo: Ministry of Education and Research.

Ogden, T. (2014). Special needs education in Norway-the past, present, and future of the field. In B. G. Cook, M. Tankersley, & T. J. Landrum (Eds.), *The Past, the Present and the Future of Special Education: Advances in Learning and Behavioral Disabilities* (Vol. 27, pp. 213-238). New York, NY: Emerald Group Publishing Ltd.

Ogden, T., Sørlie, M-A., & Hagen, K. A. (2007) Building strength through enhancing social competence in immigrant students in primary school. A pilot study. *Emotional and Behavioural Difficulties*, *12*(2), 105-117. http://dx.doi.org/10.1080/13632750701315508

Ogden, T., Sørlie, M-A., Arnesen, A., & Meek-Hansen, W. (2012). The PALS school-wide positive behavior support model in Norwegian primary schools. Implementation and evaluation. In J. Visser, H., Daniels, & T. Cole (Eds.), *Transforming Troubled Lives*. Strategies and Interventions for Children with Social, Emotional and Behavioral Difficulties, International Perspectives on Inclusive Education (Vol. 2, pp. 39-55). Bingley: Emerald Group Publishing Limited.

Olweus, D., & Limber, S. P. (2010). Bullying in school: Evaluation and dissemination of the Olweus bullying prevention program. *American Journal of Orthopsychiatry*, *80*(1), 124-134. http://dx.doi.org/10.1111/j.1939-0025.2010.01015.x

Palinkas, L. A., & Soydan, H. (2012). *Translation and Implementation of Evidence-Based Practice*. Oxford: Oxford University Press.

Rescorla, L. A., Ginzburg, S., Achenbach, T. M., Ivanova, M. Y., Almqvist, F., Begovac, I., Bilenberg, N., …Verhulst, F. (2013). Cross-informant agreement between parent-reported and adolescent self-reported problems in 25 societies. *Journal of Clinical Child & Adolescent Psychology*, *42*(2), 262-273. http://dx.doi.org/10.1080/15374416.2012.717870

Rotheram-Borus, M. J., Swendeman, D., & Chorpita, B. F. (2012). Disruptive innovations for designing and diffusing evidence-based interventions. *American Psychologist*, *67*(6), 463-476. http://dx.doi.org/10.1037/a0028180

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasiexperimental Designs for Generalized Causal Inference*. Boston/NY: Hougthon Mifflin Company.

Simonsen, B., Eber, L., Black, A. C., Sugai, G., Lewandowski, H., Sims, B., & Myers, D. (2012). Illinois statewide positive behavioral interventions and supports: Evolution and impact on student outcomes across years. *Journal of Positive Behavior Interventions*, *14*(1), 5-16. http://dx.doi.org/10.1177/1098300711412601

Skaalvik, E. M., & Skaalvik, S. (2007). Dimensions of teacher self-efficacy and relations with strain factors, perceived collective teacher efficacy, and teacher burnout. *Journal of Educational Psychology*, *99*(3), 611-625. http://dx.doi.org/10.1037/0022-0663.99.3.611

Sørlie, M-A., & Ogden, T. (2007). Immediate impacts of PALS: A school-wide multi-level programme targeting behaviour problems in elementary school. *Scandinavian Journal of Educational Research*, *51*(5), 471-492. http://dx.doi.org/10.1080/00313830701576581

Sørlie, M-A., & Ogden, T. (2014). Reducing threats to validity by design in a nonrandomized experiment of a school-wide prevention model. *International Journal of School & Educational Psychology*, *2*(4), 235-246. http://dx.doi.org/10.1080/21683603.2014.881309

Sørlie, M-A., & Ogden, T. (2015). School-Wide Positive Behavior Support-Norway. Impacts on problem behavior and classroom climate. *International Journal of School & Educational Psychology*, *00*, 1-16. http://dx.doi.org/10.1080/21683603.2015.1060912

Sørlie, M-A., Ogden, T., & Olseth, A. R. (n.d.). *Examining teacher outcomes of the school-wide positive behavior support model in Norway: Perceived efficacy and behavior management*.

Sørlie, M-A., & Nordahl, T. (1998). *Problematferd i skolen: Hovedfunn, forklaringer og pedagogiske implikasjoner*. Rapport: 12a/98. Oslo: NOVA.

Sørlie, M-A., & Ogden, T. (2014). *Mindre problematferd i grunnskolen*? *98*, 190-202. Lærervurderinger i et 10-års perspektiv . Norsk Pedagogisk Tidsskrift.

Sørlie, M-A., Ogden, T., Arnesen, A., Olseth, A. R., & Meek-Hansen, W. (2014). *Skoleledere om PALS. Spesialpedagogikk*, *1*, 17-29.

Sugai, G., & Horner, R. H. (2009). Defining and describing school-wide positive behavior support. In W. Sailor, G. Dunlap, G. Sugai, & R. H. Horner (Eds.), *Handbook of Positive Behavior Support*, pp. 307-326. New York: Springer Verlag.

Sugai, G., Horner, R. H., & Todd, A. W. (2009). Effective Behavior Support (EBS) Self-Assessment Survey (Version 3). Eugene, OR: University of Oregon, Educational and Community Supports. Retrieved July 01, 2015 from https://www.pbisapps.org/Applications/Pages/PBIS-Assessment-Surveys.aspx#sas

Waasdorp, T. E., Bradshaw, C. P., & Leaf, P. J. (2012). The impact of schoolwide positive behavioral

interventions and supports on bullying and peer rejection: A randomized controlled effectiveness trial. *Archives of Pediatrics & Adolescent Medicine*, *166*(2), 149-156. http://dx.doi.org/10.1001/archpediatrics.2011.755

Weare, K., & Nind, M. (2011). Mental health promotion and problem prevention in schools: what does the evidence say? *Health Promotion International*, *26*, (Suppl. 1), i29-i69. http://dx.doi.org/10.1093/heapro/dar075

Wendelborg, C. (2011). *Mobbing, diskriminering og uro i klasserommet. Analyse av Elevundersøkelsen 2011*. Trondheim: NTNU Samfunnsforskning, Avdeling for mangfold og inkludering.

Whear, R., Thompson-Coon, J., Boddy, K., Fordt, T., Racey, D., & Stein, K. (2012). The effect of teacher-led interventions on social and emotional behavior in primary school. A Systematic Review. *British Educational Research Journal*, *Ifirst*, 1-38.

Wilson, S. J., & Lipsey, M. W. (2007). School-based interventions for aggressive and disruptive behavior: update of a meta-analysis. *American Journal of Preventive Medicine*, *33*(Suppl. 2): S130-S143. http://dx.doi.org/10.1016/j.amepre.2007.04.011