

# Exploring Family Profiles in Explaining Heterogeneity in Parenting Program Engagement and Effectiveness

Weeland, Joyce; de Haan, Amaranta; Scott, Stephen; Seabra-Santos, Maria Joao; Webster-Stratton, C.; McGilloway, Sinead; Matthys, Walter; Gaspar, Maria; Williams, Margiad; Morch, Willy-Tore; Axberg, Ulf; Raaijmakers, Maartje; Leijten, Patty

**Journal of Family Psychology**

*DOI:*
10.1037/fam0001275

Published: 01/03/2025

Peer reviewed version

**Abstract**

Parenting programs have proven effective in reducing disruptive child behavior. However, not all families benefit equally, and, to date, we have little insight into who benefits more or less, and why. One possible solution is to explore how different potential moderators cluster together in individual families and whether such family profiles predict who benefits more or less from these programs. This study explores: (1) how family, child, and parenting risk factors for disruptive behavior cluster together in families enrolled in the popular and evidence-based Incredible Years Parenting Program using latent profile analyses; (2) how family profiles relate to covariate family characteristics; and (3) whether profiles predict program engagement (i.e., number of sessions attended by caregivers), and effectiveness of (i.e., pre-post changes in disruptive behavior). Individual participant data from six studies across four countries (Norway, the Netherlands, England, Portugal) were used, including a total sample of 772 families with children aged 2.5 to 9 years ($M$ = 5.14; $SD$ = 1.10; 58.0% boys). Families could be profiled into a low and high risk profile, which differed on most child and family (but not parenting) risk factors as well as on covariate family characteristics, such as severity of disruptive behavior. However, profile membership did not predict engagement in, or effectiveness of, the program. These findings provide useful insights into the heterogeneity in families participating in parenting programs, although there is a need for further research on how such differences may relate to differences in program effectiveness.

*Keywords*: disruptive behavior; family risk; moderation; parenting program; profiles

# Exploring Family Profiles in Explaining Heterogeneity in Parenting Program Engagement and Effectiveness

Behavioral parenting programs which incorporate parenting and disciplinary techniques for caregivers, have consistently been shown to prevent and decrease disruptive child behavior (DB) (see for reviews and meta-analyses Bausback & Bunge, 2021; Beelmann et al., 2023; Mingebach et al., 2018). However, these programs are not equally effective for all families. A synthesis of 26 meta-analyses on parenting programs showed that while there is strong evidence that, *on average,* these programs effectively reduce DB, there is considerable heterogeneity (Mingebach et al., 2018; Weber et al., 2019). Efforts to explain this heterogeneity have typically explored family characteristics that may moderate (i.e., enhance or reduce) program effects. These studies have explored, among others, family (e.g., socio-economic status, caregiver depression) and child (e.g., severity of DB) characteristics to predict which families benefit more or less (Dedousis-Wallace et al., 2020; Leijten et al., 2020; Piquero et al., 2016).

However, most moderation studies on interventions for children with DB in general (McMahon et al., 2021), and more specifically parenting programs (Leijten et al., 2020), yield inconsistent conclusions. This is mostly true for single effectiveness studies, which are often powered to assess mean program effects, and which may, therefore, be underpowered to assess moderator effects (e.g., Weeland et al., 2017). This also seems to be the case, though, for studies that have pooled data across multiple studies, and which theoretically have greater statistical power to identify moderators. Few individual family characteristics consistently moderate parenting program effects (Leijten et al., 2020; Leijten, Raaijmakers, et al., 2018).

This raises questions about the methodological and theoretical limitations of traditional approaches to moderation. First, heterogeneity in effectiveness is explained by multiple child, parent and family characteristics (Klahr & Burt, 2014; McMahon et al., 2021). Theoretical

frameworks such as the family system theory (Minuchin, 1988), family stress theory (Hill, 1959), and parental self-efficacy theory (Bandura & Adams, 1977) suggest that such characteristics are likely to interact in complex ways. Such complex interactions elude our traditional moderation studies since these studies use a variable-centered approach. In variable-centered analyses the covariation among moderator variables is modeled with the assumption that the sample is homogenous beyond the selected moderator variable (Howard & Hoffman, 2018; McMahon et al., 2021). In short, in these analyses it is assumed that families may be different on a moderator variable but the same on all other variables. One possible solution may be to assess a constellation of family characteristics that potentially moderate effectiveness of parenting programs.

Second, most traditional moderator studies focus on single moderators. Moreover, the theoretical basis for the selection of these moderators, and consequently the practical relevance of the findings, is often unclear. Although such moderator studies thus provide some insights into for whom interventions (do not) work, a more holistic approach into understanding differential effectiveness of interventions necessitates a more comprehensive selection and clustering of (theoretically) based moderator variables. An alternative and potentially more useful approach to moderation may be to select multiple family characteristics based on our *theories* on change. For example, specific clustering of caregiver risk factors may prevent changes in parenting behavior, or a specific clustering of child risk factors may maintain DB in spite of an intervention on parenting. In sum, exploring how these different characteristics cluster within families, may help us better predict which families benefit more or less and help us form hypotheses on why this may be the case.

**A Family-Centered Approach: Family Risk Profiles**

The ways in which different parenting, child and family characteristics clusters within a family may differ between families. Although families participating in parenting programs may

be similar to each other in terms of experiencing their child's behavior as disruptive, they are likely to also be different (and therefore heterogeneous), on other family characteristics. Theoretically, there may be endless numbers of combinations of characteristics and thus many subgroups of families. However, categorizing our populations into all possible subgroups leads to a high Type I error rate and low statistical power (Lanza & Rhoades, 2013). Moreover, it is unlikely that all combinations of characteristics exist in populations and are equally relevant for explaining effectiveness of interventions. Family-centered analyses (e.g., cluster analysis, latent class/profile or growth mixture modelling) that cluster families based on their shared characteristics instead of clustering (or factor-analyzing) variables used in traditional variable-centered analyses (e.g., ANOVA, regression, correlation, factor analysis) may offer a solution to this problem. A family-centered approach, in which families are grouped together into latent profiles based on their similarities with families in the same profile and differences with families in other profiles, would enable us to model complex multivariate patterns of caregiver, child and family characteristics. This has the potential to identify subgroups that may be missed using traditional moderation analyses, such as pre-defined categorical subgroups.

Pioneering studies have successfully used latent profile-analyses to assess how different risk factors cluster within families and relate to intervention effectiveness. Interestingly, the indicated family profiles often offer new hypotheses on why some families benefit more than others. For example, Dale and colleagues (2021) classified families participating in *The New Forest Parenting Program* (a home-based manualized intervention for preschoolers with ADHD) into three distinct family profiles which were mostly characterized by: parental depression, parental anxiety, and/or consistently elevated scores across risk factors (high overall family stress). These profiles predicted the extent to which families benefitted from the programs in terms of improving parenting behavior, but not child behavior (Dale et al., 2021). Specifically, families with a profile characterized by caregiver depression benefitted the least.

The authors hypothesize that caregiver depression increases the likelihood that caregivers respond to DB in a negative way and interfere their ability to implement effective parenting techniques (Dale et al., 2021).

Similarly, Pelham and colleagues (2017) found five different family risk profiles in families who participated in the well-known *Head Start* program, a comprehensive service in the US for economically disadvantaged families. The profiles were characterized by (a) relatively high income, low-risk; (b) low income, lower education, very high maternal depression, high single parenthood; (c) low income, lower education, high single parenthood, otherwise low-risk; (d) lower education, high child behavior problems, very high number of kids, high parental neglect, high maternal depression; and (e) high legal problems, very high neglect, extremely high parental mental health treatment. The authors note that the profiles differ specifically with regard to families' overall risk as well as their demographic and parental mental health risk. These profiles predicted how much families benefitted from the program in terms of decreasing DB. Specifically, families with a profile that was characterized by high overall risk, including demographic and parental mental health risk, benefited from *Head Start* but the program was not effective for families characterized by low risk or demographic risk (but otherwise low risk) (Pelham et al., 2017). The authors hypothesize that families with more risk factors (across domains), i.e., those most in need, benefit most from interventions, whereas there may be little to gain in low-risk families.

Thus, the use of a family-centered approach may help us better predict which families are likely to benefit from a specific program and help us form new hypotheses on why. This may help to increase the overall impact and cost-effectiveness of our parenting intervention strategies by, for example, specifically targeting families predicted to benefit most from parenting programs and offering alternative support to those who are likely to benefit little. At the same time, however, there are challenges to using family-centered approaches. Firstly,

many studies may not be sufficiently powered to identify less prevalent profiles and to test differential effectiveness between (possibly) multiple profiles. One way to address this, and to boost variation in selected variables and power, is to combine individual participant data from multiple studies of the same program. Secondly, the identified profiles only represent heterogeneity across the dimensions included in the model and it is important, therefore, to provide a strong theoretical basis for the selected variables used in the analyses.

**Theories on Change: Choice of Risk Factors**

The extent to which a parenting program is effective in reducing DB for an individual family, may depend on how well the content and delivery fits the specific combination of risk factors that relate to the etiology, development, or maintenance of children's DB in that specific family. Indeed, the effectiveness of programs, in general, are strongest on those family characteristics that are explicitly targeted (Leijten, Gardner, Melendez-Torres, et al., 2018; Weber et al., 2019). Most parenting programs also address topics beyond parenting behavior merely targeting DB, such as academic, persistence, social and emotional coaching of children and beyond parenting behavior, such as caregiver cognitions and emotions or self-care (e.g., Webster-Stratton, 2015). There is some evidence that the effects of these programs also reach beyond parenting behavior (Colalillo & Johnston, 2016; Feldman & Werner, 2002; McGilloway et al., 2012). At the same time, their main focus is on parenting and the key mechanism of change in DB is parenting behavior. Theoretically therefore, we may expect that parenting programs are most effective for families in which parenting challenges are more severe. Indeed, it has been shown that caregivers with high initial levels of critical, harsh, and ineffective parenting benefited most from the parenting program Incredible Years (Beauchaine et al., 2005; Reid, Webster-Stratton, & Baydar, 2004).

At the same time, families of children with DB are likely to differ from each other in terms of the risk factors that contribute to their children's DB. The risk factors that contribute

to the etiology, development and maintenance of child DB are known to be substantially heterogeneous (Beauchaine & McNulty, 2013; Bolhuis et al., 2017). We selected the factors used for the family risk profiles in the present study based on the strength of empirical evidence from reviews and cohort studies on the etiology of DB. In addition to risk factors in the domain of parenting behavior (i.e., harsh discipline, low positive parenting), two other sets of risk factors for the onset and development of DB in children have been repeatedly identified: child risk factors (specifically emotional problems and hyperactivity) and family risk factors (specifically caregiver depression, age and education) (see for reviews and cohort studies Carbonneau et al., 2022; Gutman et al., 2019; Malcolm-Smith et al., 2022; Petersen et al., 2015). Typically, children with high levels of emotional problems and/or hyperactivity and/or who grow up with younger, less educated, more depressed caregivers are at increased risk of developing DB and continuing to exhibit externalizing behaviors throughout childhood. Parenting programs do not specifically target these risk factors and may be less effective, therefore, for families in whom risk factors for DB are more child or family-related rather than parenting-related. Indeed, in the example of *The New Forest Parenting Program*, caregivers with a family profile that included high caregiver depression, benefitted least in terms of changes in parenting behavior, possibly because depression was a disruptive factor in effecting any change (Dale et al., 2021).

The effects of these risk factors on parenting program engagement and effectiveness may be explained in different ways. First, family and child risk factors may temper caregivers' readiness, willingness, and ability to properly engage in the program, thereby indirectly affecting program effectiveness (Brown et al., 2012; Chacko et al., 2016; but see for conflicting results Gross et al., 2015). Family risk factors, such as caregiver financial stress or depression, may increase barriers to participate in a parenting program (Kazdin & McWhinney, 2018; Kjøbli et al., 2013; Rostad et al., 2018); negatively affect program attendance (Chacko et al.,

2016; Duppong-Hurley et al., 2016); or caregivers' ability to implement the program material in daily life (Barnett, 2008; Tan et al., 2015). For example, families in which caregivers have lower levels of education are more likely to never attend or drop out of parenting programs after enrollment (Chacko et al., 2016). Moreover, caregivers who perceive more stress and lower quality of life report that a parenting program seems overwhelming and were shown to form poorer therapeutic alliances with practitioners (Kazdin & McWhinney, 2018; Kjøbli et al., 2013; Rostad et al., 2018), although the findings in this regard are mixed (e.g., Pereira et al., 2014; Smith et al., 2018). Overall, the weight of evidence suggests that engagement is crucial. Only through effective engagement will we be able to distinguish between those families for whom the program is less effective because they do not engage versus those who do engage, but seemingly do not benefit.

Second, if family and child risk factors are not (properly) addressed and/or persist during program participation they may limit change in parenting and/or child behavior during or initial effects on parenting and child behavior may fade over time (e.g., families falling back into old behavior patterns). For example, comorbid ADHD symptoms may evoke coercive patterns between caregiver and child (Beauchaine & McNulty, 2013); children's emotional symptoms, such as difficulties in emotional regulation, may maintain DB despite changes in parenting behavior (Scott & O'Connor, 2012); or caregiver financial stress or depression may maintain parenting problems (S. H. Goodman et al., 2020). Indeed, on average, comorbid emotional problems in children, and depression in caregivers, have been shown to be unaffected by parenting programs (for reviews and meta-analyses see Buchanan-Pascall et al., 2018; Daley et al., 2018; Leijten, Gardner, Landau, et al., 2018). Findings on ADHD symptoms are mixed (see Groenman et al., 2022; Daley et al., 2018). The evidence that these risk factors in isolation moderate program effectiveness has also been mixed (Baumel et al., 2021; Dedousis-Wallace et al., 2020; Leijten et al., 2020; Leijten, Raaijmakers, et al., 2018; Piquero

et al., 2016). However, the way they cluster within families may affect the extent to which families engage in, and benefit from, parenting programs (Griest & Forehand, 1983; Kazdin & McWhinney, 2018; Rostad et al., 2018; Scott & Dadds, 2009).

**Present Study**

The aims of this study are to: (1) explore, using latent profile analyses, how family, child, and parenting risk factors cluster together in families enrolled in the popular and evidence-based Incredible Years (IY) parenting program; (2) assess how family risk profile membership covaries with other family characteristics such as minority status, severity of and change in DB; and (3) investigate the extent to which profile membership predicts caregivers' engagement in, and the effectiveness of the IY program. The selected risk factors for inclusion in the latent profile analyses, have been identified from research conducted on the onset and development of DB and include caregiver depression, age and education; child comorbid ADHD and emotional problems; and parenting characterized by low praise and/or high harsh discipline. We hypothesize that: (a) parenting programs are most effective in terms of a decrease in children's DB in families with profiles characterized by risk factors primarily in the parenting domain, and least effective in families characterized by family and child-related factors; (b) that parenting programs are less effective in families with profiles characterized by risk in the parenting domain *as well as* by risk in one or more other domains; and (c) that families characterized by family risk factors will show the lowest level of program engagement. To maximize variation in risk factors and power, we used *integrated data analysis* by harmonizing (i.e., combining different measures of the same construct) individual participant data from all IY studies with data available on the selected family, child, and parenting risk factors. We pooled data from 6 studies. The study goals (https://doi.org/10.17605/OSF.IO/CEWY8) and study design were pre-registered (https://doi.org/10.17605/OSF.IO/HSD3W).

# Methods

## Procedure

### *Identification of Eligible Studies*.

A comprehensive literature search was performed to identify studies with data on the effectiveness of the IY parenting program. We searched for studies in the IY program in the following databases: *ERIC, PsycINFO, CINAHL, MEDLINE, Embase, Global Health, Cochrane*. The following search terms were used: *incredible years.ab OR webster-stratton.ab OR incredible years.ti OR webster-stratton.ti*. Inclusion criteria were (1) studies on the IY Parenting Program (toddler/preschool/school age version), (2) in which child behavior (i.e., disruptive behavior, hyperactivity and emotional problems), parenting behavior (i.e., the use of praise and harsh discipline), and family characteristics (i.e., caregiver age, primary caregiver depression, and education level) were assessed and for which a trial pre-registration/protocol and/or results were published in a peer reviewed journal. No restrictions were placed on the publication year. The search yielded 1,175 citations and after the removal of duplicates, abstracts of 539 citations were screened for eligibility. Of these 539 citations 383 were excluded mostly because they did not meet inclusion criteria regarding program or population and 120 were excluded because they did not measure all the target variables (flowchart in Supplement 1, Figure 1.1). In cases of doubt about eligibility, authors were contacted for additional information about the trial or data.

### *Data Collection*

Anonymized data were requested for 36 studies. Data of 23 studies were made available, of which 17 studies were excluded after screening of the data because there was no data on item-level available or harmonization of data was not possible (e.g., no available norm scores for the instruments used). Raw, individual item-level and individual participant data were supplied for six trials and checked for consistency with trial protocols and reports. The

original studies were conducted in four countries, namely England, the Netherlands, Norway, and Portugal. Information about the studies is provided in Table 1. Ethics committee approval was received for the original studies by the local institutional review board (see references Supplement 3).

***Data Harmonization***

Each study provided data reflecting the same constructs, based on the same theoretical and operational definitions. Studies mostly used standardized measures or instruments that are well-used in the field of (developmental) psychology and have been validated and normed. However, different measures or instruments were used to assess the same construct across studies and these measures were not always scaled commensurately. Before data of separate studies could be integrated, data was harmonized: recoding of variables so that constructs are scored with identical values in each study.

Our data harmonization procedures can be characterized as logical harmonization. Our procedures were based on the harmonization procedures of large British cohort studies (McElroy et al., 2020) and procedures used in previous studies on parenting programs using individual participant data from multiple trails (Leijten, Gardner, et al 2018: Leijten et al., 2020). For most constructs we selected a primary measure, based on which measure was used in the majority of the included studies. Data from studies using a different measure for this construct were converted using norm deviation scores. This way all scores of individual participants reflect how they deviate from population-specific (based on gender and age) norms.. However, for parenting risk factors this approach was not possible because no normscores for parenting behavior exist and because in two studies quantitative information about parenting behavior was collected via an interview. For this risk factor the items that theoretically fitted the same construct from two different measurements were selected. See our harmonization protocol in Supplement 2.

The assessment of harmonizability of the different scales was based on: (a) previous studies on the validity, relations between and/or comparability, and measurement invariance of the scales (reported in Supplement 2); (b) previous studies on parenting programs using similar data from multiple trials (Gardner et al., 2019; Leijten et al., 2018); and (c) discussion among co-authors of this manuscript (functioning as expert judges) (information on match between items are reported in Table S2.1, Supplement 2). To explore the validity of the harmonized scales we analyzed correlations between variables in the harmonized data across all original studies (see Tables S2.2-S2.7, Supplement 2, for correlations per study). Data harmonization is described hereunder per construct. Harmonized data from the separate studies were then merged into a single data file which was used for analyses.

## Family Risk Profile Indicators

### *Child Risk Factors at Baseline*

**Hyperactivity Symptoms.** Studies used the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997 (4 studies), the Child Behavior Checklist 6-18 years (CBCL; Achenbach, 1991) or Child Behavior Checklist 1½–5 years (Achenbach & Rescorla, 2000, 2 studies) to assess hyperactivity. Previous studies showed that both SDQ and CBCL are valid instruments for assessing hyperactivity (e.g., Riglin et al., 2021; Schmeck et al., 2001) and that the Hyperactivity scale of the SDQ is associated with the CBCL attention problem scale (Maurice-Stam et al., 2018; Theunissen, de Wolff, & Reijneveld, 2019; Vugteveen et al., 2021). The scales contain comparable items, for example “Restless, overactive, cannot stay still for long” vs. “Can't sit still/restless/hyperactive” (Table S2.1 in Supplement 2). Moreover, a study among four cohort studies showed that harmonized dimensions of emotional and hyperactivity/inattention problems are invariant across the CBCL and SDQ (Baumann et al., 2024). The SDQ was found to be invariant across different countries, however findings on

cross-country measurement invariance of the CBCL are mixed (Foley et al., 2023; Stevanovic et al., 2017).

The hyperactivity subscale of the SDQ was the most frequently used measure to assess hyperactivity (4 studies). This scale consists of 5 items measured on a 3-point scale (0='not true', 1='somewhat true', and 2='certainly true'). Two studies used the CBCL attention problems scale, respectively consisting of 10 and 5 items measured on a three-point scale (1='not true', 2='somewhat or sometimes true', 3='very or often true'). Age and gender specific population means and standard deviations were used to convert standardized CBCL sum scores into SDQ sum scores (Supplement 2).

**Emotional Problems.** Studies used the SDQ (4 studies) or CBCL (2 studies) to measure emotional problems. Previous studies showed that both SDQ and CBCL are valid instruments for assessing emotional problems (e.g., Ferdinand, 2008; Goodman et al., 2003) and that the emotional symptoms scale of the SDQ is associated with the CBCL anxious-depressed scale (Maurice-Stam et al., 2018; Theunissen, de Wolff, & Reijneveld, 2019; Vugteveen et al., 2021). The scales contain comparable items, for example "Often unhappy, depressed or tearful" vs. "Unhappy, sad, or depressed" (Table S2.1). The emotional symptoms subscale of SDQ (Goodman, 1997) was most often used to measure emotional problems (4 studies) and was therefore chosen as the primary measure. The emotional symptoms scale consists of 5 items measured on a 3-point scale (0='not true', 1='somewhat true', and 2='certainly true'). Two studies used the CBCL 4-18 years or CBCL ½–5 years anxious/depressed scale consisting of respectively 13 items and 8 items measured on a three-point scale (1='not true', 2='somewhat or sometimes true', 3='very or often true'). Age and gender specific population means and standard deviations were used to convert standardized CBCL sum scores into SDQ sum scores (Supplement 2).

***Family Risk Factors at Baseline***

**Caregiver Education.** Primary caregiver education was used as a marker of socio-economic status (SES). Caregivers' formal education level was classified using International Standard Classification of Education (ISCED) (UNESCO Institute of Statistics, 2011) and collapsed into three categories: 1=secondary education or less; 2=post-secondary education but not university; 3=bachelor degree or higher. Low level of caregiver education is seen as a risk factor for DB.

**Caregiver Age.** Caregivers' age at the time of the birth of the target child (i.e., for which they receive IY) was calculated by subtracting the age of the target child from the age of the primary caregiver. Younger age at time of children's birth is seen as a risk factor for DB.

**Caregiver Depression.** Depression was assessed with The Beck Depression Inventory (BDI; Beck & Beamesderfer, 1974, 2 studies), the General Health (GHQ-12**;** Goldberg et al., 1997, 2 studies) or the Symptom Checklist (SCL; Arrindell & Ettema, 2005, 2 studies). Previous studies showed that all three instruments are valid measures to screen for depression in the general population (e.g., Aalto, 2012; Lasa et al., 2000; Lundin et al., 2015) and that scores on the BDI and GHQ were associated (Aalto et al., 2012), as well as the BDI and SCL depression scale (Koeter, 1992). The instruments contain comparable items, such as "I feel I am a complete failure as a person" vs. "Been thinking of yourself as a worthless person" vs. "Feelings of worthlessness" (Table S2.1). The BDI was found to be invariant across different countries (Dere et al., 2015; Nuevo et al., 2009). However, information about measurement invariance of the other scales is lacking.

The BDI was used as the primary measure for caregiver depression. This scale consists of 21 items measured on a 4-point scale (0=*not at all* to 3=*severely*). The 12 GHQ items are measured on a 4-point scale (ranging from "*better/healthier than normal"* option to a "*much worse/more than usual"*). The Depression subscale of the SCL consists of 16 items measured on a 5-point Likert scale (0=*not at all* to 4=*extremely*). Population means and standard

deviations were used to convert standardized GHQ and SCL sum scores into BDI sum scores (Supplement 2). For one study, caregiver depression was only assessed in families who were allocated to the intervention condition.

### *Parenting Risk Factors at Baseline*

Parenting risk factors consist of both the presence of negative parenting behavior and the absence of positive parenting behavior. Based on a previous parenting profile analysis caregivers' low self-reported use of praise and high self-reported use of harsh discipline were used as markers for parenting risk (Weeland et al., 2022). In most studies (5 studies) parenting behavior was measured using the Parenting Practice Interview questionnaire (PPI; Webster-Stratton, 2001). The PPI was therefore used as primary measure. In two studies parenting behavior was measured with a Parenting Interview (Dowdney et al., 1985). Items from the questionnaire and interview questions which theoretically measured the same construct were matched. Almost identical questions were used across instruments, namely: for praise: "How often do you praise or compliment your child when your child behaves well or does a good job?" vs. "How many times per day did you praise your child for doing something you asked them or doing something well?"; for harsh discipline: "How often do you do each of the following things when your child misbehaves? Slap or hit your child (but not spanking)" vs. "Thinking about last week, how many times did you give your child a tap or smack if he/she misbehaved". One questionnaire item was selected to measure praise and six to measure harsh discipline ($\alpha = .86$). Response scales were harmonized to reflect the response scale used most frequently, which was a 7-point Likert scale. Scores from the interview were therefore converted from a 5-point to a 7-point Likert scale (Supplement 2).

### *Distal Outcome: Disruptive Child Behavior*

Changes in DB between pre- and post-test was used as the distal outcome (i.e., the observed variable that is predicted by the latent categorical variable) and assessed using the Eyberg Child

Behavior Inventory—Intensity Scale (ECBI; Eyberg & Pincus, 1999, 5 studies) or SDQ 'Conduct Problems' sub-scale (1 study). Both instruments have been shown to be valid screening measures for conduct problems in children (e.g., Abrahamse et al., 2015; Goodman et al., 2003) and previous studies have shown that the ECBI intensity and SDQ conduct problem scales are associated (Abrahamse et al., 2015). The instruments contain comparable items, such as "Often loses temper" vs. "Has temper tantrums". The SDQ was found to be invariant across different countries (Foley et al., 2023; Stevanovic et al., 2017), however studies on measurement invariance of the ECBI are lacking.

The most frequently used measure for caregiver-reported DB was the ECBI, consisting of 36 items measured on a 7-point Likert scale (1=*never* to 7=*always*). For one trial, scores on SDQ 'Conduct Problems' sub-scale were therefore converted into the 5-item ECBI Intensity Scale. The SDQ Conduct Problem sub-scale age- and gender- specific population means and standard deviations (Maurice-Stam et al., 2018) were used to calculate standardized sum scores, which were converted to ECBI sum scores (Weeland et al., 2018) (Supplement 2).

**Participants**

The total integrated sample from the studies included data on 772 families (Table 1). Children were aged 2.5 to 9 years old ($M$ = 5.14; $SD$ = 1.10; 58.0% boys). Most caregivers were female (96.9%), approximately one quarter of caregivers were single (26.1%), half (54.8%) reported to be part of an ethnic minority, and 43.1% were educated to secondary level or lower. If multiple caregivers from one family were included in the study, we used data from one caregiver. All participants gave informed consent for participation in the original study (Supplement 3).

**Intervention**

In each of the studies, families were allocated to an intervention (57.0%) or control (43.0%) condition. Families in the intervention condition were offered the Incredible Years

(IY) parenting program (Webster-Stratton 2015). IY is a group behavioral parent training program which starts with the focus on positive parenting strategies such as child-led play, social en emotional coaching, praise, and incentives, before discussing effective limit setting, ignoring unwanted behavior, and finally, consequences of DB, problem-solving skills, and time-out strategies. IY uses a collaborative setting, in which group leaders establish themselves as facilitators rather than as experts (Webster-Stratton, 2012). Group leaders encourage caregivers to solve problems and to help one another in this regard to ensure maintenance of the intervention effects. The IY program is specifically designed to attune to variations in, amongst others, caregivers' cognitions and learning skills In all studies, at least one of the two group leaders in each group, was a certified IY group leader, or was undergoing certification process (Table 3.1, Supplement 3). The number of sessions offered to participants in the included studies ranged from 12 to 18, depending on the version of the program. Caregivers across studies attended, on average, 73% of the sessions offered to them (range from 0 (4.5%) to all (19.0%)). Families in the control condition were either on a waitlist for IY (3 studies), received a minimal intervention through the study (e.g., a phone helpline; 1 study), or received no active intervention through the study (2 studies).

***General Effects of the IY Intervention on Disruptive Child Behavior***.

The treatment condition predicted DB at T2; that is, families who received IY reported on average less DB at T2 than those families who did not (small effect, Cohen's $d = .27$). See Supplement 4. We calculated individual reliable change indexes (RCIs) to evaluate whether a change within a child's DB is greater than a difference that could have occurred due to random measurement error (i.e., reliable) and is clinically meaningful (Jacobson et al., 1999). For the calculation of these index we used: the ECBI means and standard deviation of DB at T1 and T2 of the sample; the ECBI means and standard deviation of a well-functioning normal population; (Weeland et al., 2018); and the ECBI test-retest reliability (Abrahamse et al., 2015). One fifth

(20.4%) of children of families in the intervention condition showed recovery or clinically significant improvement (vs. 7.3% in the control group). Thus, although, on average, the intervention had a significant effect on DB, there was large heterogeneity in effect size between individual families. This emphasizes the importance of understanding heterogeneity in intervention effectiveness.

**Main Analysis**

Data were analyzed using M*plus* (Muthén & Muthén, 2017). A three-step latent profile analysis (i.e., a class analysis with continuous and categorical data) with distal outcomes was used to identify family risk profiles and to explore whether profile membership predicted parenting program engagement and effectiveness (Asparouhov & Muthen, 2013). Latent profile analyses are commonly used for identifying subpopulations within a population based on a certain set of variables and is a statistically sophisticated technique to identify finite subgroups that are not directly observable (Lanza & Rhoades, 2013; Ferguson, Moore, & Hull, 2020).

***Step 1: Determination of Latent Profiles***

In step one, the number of latent family profiles is determined without the distal outcomes that will be part of the secondary model. The profile indicators were based on pre-intervention data on parenting (praise and harsh discipline), family (caregiver age, education and depression), and child (hyperactivity symptoms and emotional problems) risk factors. Missing data on these profile indicators ranged from 6.7% (for caregiver education) to 19.8% (caregiver depression). When compared to complete cases, those with missing data on profile indicators had younger children, were more likely to be allocated to the control group and less likely to identify themselves as part of an ethnic minority. To address missing data patterns and maximize the number of families we could classify in a profile (i.e., prevent listwise deletion during profile allocation) we used multiple imputation, simulating random draws from the

posterior distribution of the missing scores. These scores were generated using all profile indicators, and the indicators of missing data. We used 50 imputed datasets.

Latent profile analyses from 1- to 5-profile solutions were ran sequentially on the 50 imputed datasets and were evaluated based on: (i) three fit indices (Bayesian information criterion (BIC), Akaike information criterion (AIC), and the Lo-Mendell-Rubin adjusted (LRT) test), (ii) entropy (i.e., estimate of the probability that each participant is in each of the classes) and mean class probabilities, (iii) profile size (i.e., at least 5% of participants in each profile), and (iv) theoretical plausibility (i.e., whether subgroups seem theoretically meaningful). After the selection of the best profile solution, we reran the class analysis on one of the imputed datasets, fixing the profile indicators based on the output in step 1.

***Step 2: Allocation to Latent Family Profile and Determining Measurement Error***

In step 2, families were allocated to one of the profiles based on the profile solutions selected in step one. We based allocation on the profile which has the highest probability for the individual family. In this step we also retrieved the measurement error for the most likely profile (i.e., Logits for the Classification Probabilities the Most Likely Latent Class Membership), which will be used in step 3. Profile membership and probability scores were merged with the larger dataset.

***Step 3: Relation between Profiles and Distal Outcomes***

In step 3, the relationship between profile membership and distal outcomes (i.e., predicted parenting program engagement and effectiveness) were assessed while accounting for classification errors determined in step two. In all steps, we report 95% confidence intervals (CI) of means or paths for each profile (non-overlapping CI indicating significant differences).

We did this in two phases: We first explored whether and how families allocated to different profiles differed from each other on the profile indicators as well as other child and family characteristics (other than those used to identify the profiles), specifically experimental

condition (percentage allocated to IY), minority status (percentage minority status), family composition (percentage single caregivers), and DB at T1. Second, we tested whether families allocated to different profiles showed different engagement in (i.e., percentage of attended IY sessions), and differentially benefited from, IY (i.e., path from condition to DB at T2 and path from condition to reliable change score). In this step we controlled for the original study of origin, gender, and DB stability (i.e., relation between DB at T1 and T2). Data is not available, but study materials and syntax codes are available (https://osf.io/cewy8/). The syntax codes for step 1 and 3 is found in Supplement 6.

## Results

### Step 1. Determination of Latent Profiles

The sequential profiles analyses showed that AIC and BIC decreased when the number of profiles increased, indicating increased model fit with more profiles. The LRT test was only significant for the step from 1 to 2 profiles and from 3 to 4 profiles, but not so from 2 to 3 profiles or from 4 to 5 profiles. Entropy increased with rising numbers, but class probabilities decreased. For the two-profile solution, the smallest profile consisted of 45.7% of participants, but this percentage dropped under 5% from the three-profile solution onwards (Table 2). We decided on a two-profile solution since: (i) the fit of a two-profile solution was significantly better than a one-profile solution; (ii) entropy was satisfactory (.82) and class probabilities were high (above .92 for both profiles), indicating that participants could be allocated to a certain profile with high probability; (iii) the sizes of both profiles were substantial; and (iv) this solution seems theoretically plausible since at face value, the profiles indicated that most risk factors were higher in one profile than the other.

### Step 2. Allocation to Latent Family Profile and Determining Measurement Error

Families in the two profiles scored significantly different on all child risk factors and on most family risk factors (except for caregiver depression) but did not differ with regard to

parenting-related factors (Table 3). Families in profile 1 reported significantly more child emotional problems and hyperactivity symptoms, and caregivers had lower levels of education and were younger at the time the target child was born, when compared to families in profile 2. Thus, we labelled profile 1 as “high risk” and profile 2 as “low risk” profile (i.e., relative to the other profile) (Figure S5.1, Supplement 5). We explored whether these profiles differed on child- and family other than the profile allocators (Table 4). The profiles did not differ on the percentage of families allocated to the intervention condition, but they did differ on minority status, percentage of single caregivers and DB at T1. Families in the high risk profile were more often majority and single caregiver-families of which the child scored higher on DB at pretest than families in the low risk profile.

**Step 3. Relation between Profiles and Distal outcomes**

The sample size in the secondary model was lower due to missing data in the distal outcome, but distribution of families across profiles remained similar (entropy = .658, lowest class probability = .923). Results of the model showed that, (a) based on the percentage of IY sessions attended by caregivers, engagement significantly differed between profiles and (b) that IY had a significant effect on DB at post-test in both profiles. High risk families allocated to the intervention group on average attended less session than low risk families in the intervention group. In both profiles, caregivers who were allocated to the intervention group on average scored their child lower on DB at posttest compared to caregivers in the control condition in both profiles (Table 5). Importantly, this effect of IY on DB problems did not significantly differ between the profiles (for βs, 95% CI of βs and Cohen’s *d*, see Table 5).

## Discussion

Parenting programs to prevent and decrease disruptive child behavior (DB) are not equally effective for all families. In the present study, we (1) assessed family profiles based on theoretical and empirical research on risk factors for the onset, development, and maintenance

of DB; (2) explored whether and how these profiles differed on covariate family characteristic and (changes in) child behavior, and (3) analyzed whether and how latent family risk profiles predicted differential engagement in, and effectiveness of, the Incredible Years Parenting Program (IY). Our results indicated that families could be allocated into a low- versus high risk profile (aim 1). Families in low- and high risk profiles differed on covariate characteristics, such as being a single caregiver household and severity of DB (aim 2), and they differed in their engagement, but not in the extend to which they benefitted from IY (aim 3).

The identified profiles varied in terms of most the pre-defined child (i.e., hyperactivity symptoms and emotional problems) and family (i.e., caregiver education and age at time of birth of the child) (but not parenting) risk factors for DB and are therefore labelled low and high risk profiles. These profiles may be in line with theoretical frameworks explaining how risk factors within a family interact across domains, and in some cases, maintain or intensify each other (Barnett, 2008; Cowan et al., 1998; Kreppner & Lerner, 1989; Masarik & Conger, 2017) and empirical research showing that multiple child and family risk factors tend to cluster within families (Beauchaine & McNulty, 2013; Mullola et al., 2021). However, the identified profiles did not differ on parenting risk factors (i.e., low parental praise and high harsh punishment)or caregiver depression. This in line with previous profile studies showing that caregiver depression or parenting problems do not necessarily cluster together with other risk factors (see Pelham et al., 2017).

One possible explanation may be that since the study relies on caregivers' self-reported data, the caregiver may not perceive themselves differently in, for example, parenting risk, in reality, differences may exist. Self-reported data may reflect differences in caregivers' awareness of their own behavior, or processes underlying how caregivers report on their own and their child's functioning, such as feelings of self-efficacy, attributions about child behavior or (dis)stress (Herbers et al., 2017). In the present study however, caregiver depression was

high (>20% of caregivers across profiles scored in the clinical range) compared to the average prevalence across 27 European countries (6.38%, Van der Velden et al., 2021). This indicates that in many families of children showing DB, caregiver wellbeing is affected. This could also indicate a selection effect. Although findings are mixed, there are studies showing that caregivers who perceive parenting stress are more likely to engage in an intervention (Pereira et al., 2014; Smith et al., 2018).

Regarding covariate family characteristics, families with a high risk profile were more often single caregiver families, with an older child who showed more severe DB before the intervention compared to families with a low risk profile. The largest differences between profiles on the pre-defined risk factors seemed to be on caregiver age and education: most caregivers who became a parent in their teens and caregivers with lower levels of education were allocated to the high risk profile. The clustering of pre-selected risk factors for DB families allocated to the high risk profile, and their relation to other family characteristics, could indicate increased stress and lack of social support in these families (Bull & Mittelmark, 2009; Liang et al., 2019), which may show that those in the high risk profile are most in need of support. Overall, the identified profiles may show that specifically demographic characteristics (i.e., low SES, low caregiver age, single caregivers) distinguish between different families attending parenting programs.

Families allocated to the high risk profile, who received IY, on average attended less sessions than those allocated to the low risk profile. This is in line with previous studies showing that, families with problems in multiple domains more frequently drop out if parenting programs and attend less sessions, possibly because the program seems overwhelming or because they experience more barriers to attend sessions (Kazdin & McWhinney, 2018; Kjøbli et al., 2013; Rostad et al., 2018). Importantly, we did not find evidence that these families benefitted less from the program directly after the intervention. This contradicts our expectation

that family profiles characterized by severe parenting risk would benefit most and profiles characterized by high caregiver and child risk would benefit least, since parenting programs are not designed to treat these problems. However, it is in line with results of previous variable-centered analyses on differential effectiveness of parenting programs showing that risk factors for DB such as comorbid ADHD and emotional problems or caregiver depression in isolation do not reduce effects (Baydar et al., 2003; Leijten et al., 2020). It may be that positive changes in parenting lead to positive changes in child behavior, also in the presence of other risk factors. Parenting was indeed found to mediate the effects of cumulative risk factors on later child externalizing problems (Gach et al., 2018).

The IY program is a collaborative program specifically designed to attune to variations in, amongst others, caregivers' cognitions and learning skills. Group leaders may thus tailor program content to specific characteristics and needs of families. This means that IY may benefit different families in different ways: for some this may occur through changes in parenting or child behavior, while for others, it may relate to an increased understanding of child behavior, changes in caregiver cognitions or emotions, or increased social support through interactions with other caregivers (Feldman & Werner, 2002; Forehand et al., 2014). Even in the absence of changes in child behavior, caregivers may experience parenting programs as beneficial to them, for example due to small but important changes in their perceptions of their child's behavior, because they feel less anxious or more supported (McKay, Kennedy, & Young, 2021). One way to further increase our understanding of differential effects of parenting programs such as IY may be by exploring different and multiple outcomes of parenting programs, such as caregiver cognitions, stress, experienced support, or understanding of child behavior (development).

Our study has both important strengths and limitations. First, we used a theory-driven, family centered approach to heterogeneity in intervention effects aiming to elucidate different

processes that exist within individual families. Latent profile analysis is a sophisticated approach to moderation, that, compared to more traditional approaches, reduces type I error rates and improves power to assess higher order interactions, (Lanza & Rhoades, 2013). This enabled us to model how multiple potential moderators naturally cluster together in families and affect program engagement and effectiveness and to test complex hypotheses on why some families may benefit more than others from parenting programs. At the same time, because profile allocation is based on probabilities, there is always some uncertainty since the 'true' profile membership is unknown (Lanza & Rhoades, 2013). Replication and further validation of family profiles, for example in data on other parenting programs, is needed.

Second, we pre-selected risk factors based on theory and our current understanding of change during a parenting intervention, and how this relates to the risk factors involved in the maintenance of DB. This is important to enhance our understanding of *why* some families benefit more or less. However, many more risk factors for DB have been identified and may be important factors in the effectiveness of our intervention strategies, albeit it is not possible to include all of these in any one study. Children with DB are very heterogenous and there may be subpopulations that need a more personalized approach. Future research could explore profiles based on children's behavioral symptoms, such as callous unemotional traits (Perlstein et al., 2023), irritability (Bolhuis et al., 2017), pro-active aggression or anxiety (Rosa-Justicia et al., 2022). In general, there is a need for more (rigorous) research on such personalized procedures and adapted or personalized interventions for children with DB (review by Lane et al., 2023).

Third, the use of pooled individual participant data from multiple studies resulted in a larger and more diverse sample than is available in single studies and increased power to assess family risk profiles as a moderator of intervention effectiveness. Our studies illustrate the possibilities of international collaboration in (re)using existing data. Including families from different countries may have both advantages (e.g., diverse sample) and disadvantages (e.g.,

different norms for behaviors, see Weeland et al., 2018). The use of existing data also comes with limitations. Specifically, integration of data across studies can be challenging. In our study, different instruments were used across studies to assess the same construct. We opted for a logical approach to data harmonization. We harmonized scales from different instruments based on the clinical conceptualizations of the problems/disorders they assess, and for which reliability and validity has been studied. For each individual participant, across all original studies, the harmonized score reflects how that individual participant deviates from the norm on that specific scale. We feel this approach strengthens the clinical relevance of our approach and of the study results. It enables to assess the theoretical and clinical relevance of the found profiles and enables translation to clinical practice. For example, because we can identify families that may fall within a certain profile in future practice, and this may help us personalize our prevention and intervention strategies. However, this approach also has important limitations since it assumes the selected scales and items within the scales measure a construct equivalently across the studies. This may not always be the case. Similarly named scales do not necessarily measure the same construct (i.e., the jingle fallacy, Weidman et al., 2017). Moreover, even identical items may function differently across studies, due to -for example- regional or cultural differences in interpretation or placement of the items within study-specific test batteries (Holland & Dorans, 2006; Bauer & Hussong, 2019). Findings on measurement invariance for often used instruments such as the SDQ, CBCL, or BDI, are mixed (Foley et al., 2023; Neuvo et al., 2009; Stevanovic et al., 2017). There is a strong need for more cross-country measurement invariance studies (Stevanovic et al., 2017).

Moreover, the use of existing data may limit what can be assessed. For example, because not all studies included follow-up data (e.g., due to a waiting list design), we only assessed immediate effects of parenting programs on child behavior. Therefore, we cannot rule out that family risk profiles may predict the extent to which intervention effects are sustained.

Limited knowledge about long-term parenting programs effects, and especially about differential long-term effects is a severe limitation of the field. Families with high risk profiles may be at increased risk for problems to return, maintain or even intensify over time. In general, possibilities for the integration of individual participant data would benefit from standard measures and measure procedures across studies.

Fourth, we focus on both engagement and effectiveness. This is important in terms of distinguishing between families for which the program is less effective because they do not engage, and those who do engage but seemingly do not benefit**.** Our measure of engagement was limited to the number of sessions caregivers attended. For future studies it may be important to structurally include different measures of program fidelity and engagement, such as homework completion or therapeutic alliance (e.g., having shared goals and the bond between caregiver and group leader).

## Clinical implications

The results reported here may help to further strengthen the effectiveness of parenting programs, such as IY, in different ways, despite the fact that the identified profiles did not predict differential effectiveness. First, it is important to note that the low risk profile is considered low risk *relative to* the high risk profile but is not necessarily low risk when compared to the general population. Although approximately half of the families in this study were characterized as low risk, substantial proportions nevertheless showed comorbid children's emotional problems (31%), ADHD symptoms (33%) and caregiver depression (22%). The fact that so many families in these samples - across profiles - reported problems in multiple domains, may be important for clinical practice; these families are clearly dealing with challenges beyond their child's DB. It may be important to address these more explicitly both before and during the program.

Discussing a families' specific circumstances, challenges and their children's behavioral symptoms before they enroll in an intervention may create an opportunity to increase engagement and boost program effectiveness. It has been shown that caregivers' who feel their family's situation was assessed prior to enrollment in a parenting program perceived this program as more beneficial, compared to caregivers who felt their individual issues were never considered (McKay et al., 2020). During the program it may be important to address family issues beyond parenting, such as issues with caregiver wellbeing, partner or co-parenting relations that may limit change in parenting and child behavior. In many parenting programs, including IY, ways to do this are already incorporated in the program. Although evidence for the effectiveness of including content beyond parenting in parenting programs is mixed, a recent meta-analysis showed that for families with high clinical scores of DB, albeit not in prevention settings, the inclusion of techniques aimed at improving caregiver well-being enhanced overall program effectiveness (Leijten et al., 2019). Addressing this may have consequences for the needed intensity and duration of the program and qualifications and experience of group leaders. Moreover, for some of these issues it may be important for all caregivers to participate in the program (Weeland et al., 2021). Currently, in many cases one caregiver (mostly mothers) participates.

Second, similar to data on this from other studies (see review by Chacko et al., 2016), families on average attended approximately three-quarters of the program sessions. This is important since research has suggested that engagement is related to program outcomes (Boggs et al., 2008; Patterson & Chamberlain, 1994). There thus seems room for improvement. Group leaders could have a significant role here by, for example by offering 'make up' sessions and tailoring for factors that may relate to families dropping out (e.g. culture, family context and child developmental level). Indeed, the IY program offers group leaders ways to do this. This may be specifically important for families in vulnerable circumstances, such as those allocated

to the high risk profile. These families attended less sessions (i.e., 62%) than average (i.e., 73%) and significantly less session compared to families in the low risk profile (i.e., 86%). Future research could also explore (additional) ways to boost caregivers' engagement in parenting programs to investigate questions such as: whether discussing family problems beyond child behavior before the start of the program reduces barriers to participate; whether caregiver engagement is greater when comorbid problems are addressed first (i.e., stepped care); and whether this, in turn, enhances effectiveness (Weeland et al., 2021).

## Conclusion

Finding new ways to further strengthen the impact and cost-effectiveness of our prevention and intervention strategies may be the most significant upcoming challenge in intervention research. Our study confirms that IY is effective in reducing DB, but that the heterogeneity in effectiveness is large and difficult to predict. Although families participating in parenting programs are similar, in that they experience difficulties with their children's behavior, they also differ in many ways. In this study these differences were related to the severity and development of child behavior problems, but these factors did not predict the extent to which families engaged in, and benefitted from, the parenting program. The results of our profile analyses are important in helping us to better understand heterogeneity in families participating in parenting programs, but more work is needed to enable us to better predict and understand heterogeneity in effectiveness.

## References

Aalto, A. M., Elovainio, M., Kivimäki, M., Uutela, A., & Pirkola, S. (2012). The Beck Depression Inventory and General Health Questionnaire as measures of depression in the general population: a validation study using the Composite International Diagnostic Interview as the gold standard. *Psychiatry Research*, *197*(1-2), 163-171. https://doi.org/10.1016/j.psychres.2011.09.008

Abrahamse, M. E., Junger, M., Leijten, P. H. O., Lindeboom, R., Boer, F., & Lindauer, R. J. L. (2015). Psychometric Properties of the Dutch Eyberg Child Behavior Inventory (ECBI) in a Community Sample and a Multi-Ethnic Clinical Sample. *Journal of Psychopathology and Behavioral Assessment*, *37*(4), 679–691. https://doi.org/10.1007/s10862-015-9482-1

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA preschool forms and profiles.* Burlington, VT: University of Vermont, Research Centre for Children, Youth & Families.

Arias-de la Torre, J., Vilagut, G., Ronaldson, A., Serrano-Blanco, A., Martín, V., Peters, M., ... & Alonso, J. (2021). Prevalence and variability of current depressive disorder in 27 European countries: a population-based study. *The Lancet Public Health*, *6*(10), e729-e738. https://doi.org/10.1016/S2468-2667(21)00047-5

Arrindell, W. A., & Ettema, J. H. M. (2005). *SCL-90 Handleiding bij een multidimensionale psychopathologie-indicator [SCL-90: Manual for a multidimensional indicator of psychopathology].* Pearson.

Asparouhov, T., & Muthen, B. (2013). *Auxiliary variables in mixture modeling: 3-step approaches using M*plus (pp. 1–48). Los Angeles, CA: Muthén & Muthén. Retrieved from http://www.statmodel.com/examples/webnotes/webnote15.pdf.

Bandura, A., & Adams, N. E. (1977). Analysis of self-efficacy theory of behavioral change. *Cognitive Therapy and Research*, *1*(4), 287–310. https://doi.org/10.1007/BF01663995

Barnett, M. A. (2008). Economic disadvantage in complex family systems: Expansion of family stress models. *Clinical Child and Family Psychology Review*, *11*(3), 145–161. https://doi.org/10.1007/s10567-008-0034-z

Bauer, D. J., & Hussong, A. M. (2009). Psychometric approaches for developing commensurate measures across independent studies: traditional and new models. *Psychological methods, 14(2*), 101. https://doi.org/10.1037/a0015583

Baumel, A., Mathur, N., Pawar, A., & Muench, F. (2021). Psychosocial Interventions for Children With Externalized Behavior Problems: An Updated Meta-analysis of Moderator Effects. *Journal of Child and Family Studies*, *30*, 65–86. https://doi.org/10.1007/s10826-020-01863-6

Bausback, K. B., & Bunge, E. L. (2021). Meta-Analysis of Parent Training Programs Utilizing Behavior Intervention Technologies. *Social Sciences, 10,* 367. https://doi.org/10.3390/SOCSCI10100367

Baydar, N., Reid, M. J., & Webster-Stratton, C. (2003). The Role of Mental Health Factors and Program Engagement in the Effectiveness of a Preventive Parenting Program for Head Start Mothers. *Child Development*, *74*(5), 1433–1453. https://doi.org/10.1111/1467-8624.00616

Beauchaine, T. P., & McNulty, T. (2013). Comorbidities and continuities as ontogenic processes: Toward a developmental spectrum model of externalizing psychopathology. *Development and Psychopathology*, *25*(4pt2), 1505–1528. https://doi.org/10.1017/S0954579413000746

Beauchaine, T. P., Webster-Stratton, C., & Reid, M. J. (2005). Mediators, moderators, and predictors of 1-year outcomes among children treated for early-onset conduct problems: A latent growth curve analysis. *Journal of Consulting and Clinical Psychology, 73*, 371–388. https://doi.org/10.1037/0022-006X.73.3.371

Beck, A. T., & Beamesderfer, A. (1974). Assessment of depression: the depression inventory. *Modern Problems of Pharmacopsychiatry*, *7*(0), 151–169. https://doi.org/10.1159/000395074

Beelmann, A., Arnold, L. S., & Hercher, J. (2023). Parent training programs for preventing and treating antisocial behavior in children and adolescents: A comprehensive meta-analysis

of international studies. *Aggression and Violent Behavior*, *68*, 101798. https://doi.org/10.1016/J.AVB.2022.101798

Bolhuis, K., Lubke, G. H., van der Ende, J., Bartels, M., van Beijsterveldt, C. E. M., Lichtenstein, P., Larsson, H., Jaddoe, V. W. V., Kushner, S. A., Verhulst, F. C., Boomsma, D. I., & Tiemeier, H. (2017). Disentangling Heterogeneity of Childhood Disruptive Behavior Problems Into Dimensions and Subgroups. *Journal of the American Academy of Child & Adolescent Psychiatry*, *56*(8), 678–686. https://doi.org/10.1016/J.JAAC.2017.05.019

Brown, L. D., Goslin, M. C., & Feinberg, M. E. (2012). Relating Engagement to Outcomes in Prevention: The Case of a Parenting Program for Couples. *American Journal of Community Psychology*, *50*(1–2), 17–25. https://doi.org/10.1007/S10464-011-9467-5/TABLES/2

Buchanan-Pascall, S., Gray, K. M., Gordon, M., & Melvin, G. A. (2018). Systematic Review and Meta-analysis of Parent Group Interventions for Primary School Children Aged 4–12 Years with Externalizing and/or Internalizing Problems. *Child Psychiatry and Human Development*, *49*(2), 244–267. https://doi.org/10.1007/S10578-017-0745-9/METRICS

Bull, T., & Mittelmark, M. B. (2009). Work life and mental wellbeing of single and non-single working mothers in Scandinavia. *Scandinavian Journal of Public Health*, *37*(6), 562–568. https://doi.org/10.1177/1403494809340494

Carbonneau, R., Vitaro, F., Brendgen, M., Boivin, M., & Tremblay, R. E. (2022). Early Risk Factors Associated With Preschool Developmental Patterns of Single and Co-Occurrent Disruptive Behaviors in a Population Sample. *Developmental Psychology*, *58*(3), 438–452. https://doi.org/10.1037/DEV0001295

Chacko, A., Jensen, S. A., Lowry, L. S., Cornwell, M., Chimklis, A., Chan, E., Lee, D., & Pulgarin, B. (2016). Engagement in Behavioral Parent Training: Review of the Literature

and Implications for Practice. *Clinical Child and Family Psychology Review*, *19,* 204–215. https://doi.org/10.1007/s10567-016-0205-2

Colalillo, S., & Johnston, C. (2016). Parenting cognition and affective outcomes following parent management training: A systematic review. *Clinical Child and Family Psychology Review*, *19*(3), 216–235. https://doi.org/10.1007/s10567-016-0208-z

Cowan, P. A., Powell, D., & Cowan, C. P. (1998). Parenting interventions: A family systems perspective. In W. Damon, I. E. Sigel, & K. A. Renninger (Eds.), *Handbook of child psychology: Child psychology in practice* (pp. 3–72). John Wiley & Sons Inc.

Dale, C., Parent, J., Forehand, R., DiMarzio, K., Sonuga-Barke, E., Long, N., & Abikoff, H. B. (2021). Behavioral Parent Training for Preschool ADHD: Family-Centered Profiles Predict Changes in Parenting and Child Outcomes. *Journal of Clinical Child and Adolescent Psychology*, 1–14. https://doi.org/10.1080/15374416.2020.1867987

Daley, D., Oord, S. Van Der, Ferrin, M., Cortese, S., Danckaerts, M., Doepfner, M., Hoofdakker, B. J. Van den, Coghill, D., Thompson, M., Asherson, P., Banaschewski, T., Brandeis, D., Buitelaar, J., Dittmann, R. W., Hollis, C., Holtmann, M., Konofal, E., Lecendreux, M., Rothenberger, A., … Sonuga-Barke, E. J. (2018). Practitioner Review: Current best practice in the use of parent training and other behavioural interventions in the treatment of children and adolescents with attention deficit hyperactivity disorder. *Journal of Child Psychology and Psychiatry*, *59*(9), 932–947. https://doi.org/10.1111/JCPP.12825

Dedousis-Wallace, A., Drysdale, S. A., McAloon, J., & Ollendick, T. H. (2020). Parental and Familial Predictors and Moderators of Parent Management Treatment Programs for Conduct Problems in Youth. *Clinical Child and Family Psychology Review*, 1–28. https://doi.org/10.1007/s10567-020-00330-4

Dowdney, L., Skuse, D., Rutter, M., Quinton, D., & Azek, D. (1985). The Nature and Qualities of Parenting Provided by Women Raised in Institutions. *Journal of Child Psychology and Psychiatry*, *26*(4), 599–625. https://doi.org/10.1111/j.1469-7610.1985.tb01644.x

Duppong-Hurley, K., Hoffman, S., Barnes, B., & Oats, R. (2016). Perspectives on Engagement Barriers and Alternative Delivery Formats from Non-completers of a Community-Run Parenting Program. *Journal of Child and Family Studies*, *25*(2), 545–552. https://doi.org/10.1007/s10826-015-0253-0

Eyberg, S. M., & Pincus, D. (1999). Eyberg child behavior inventory and Sutter-Eyberg student behavior inventory-revised: Professional manual. Psychological Assessment Resources.

Feldman, M. A., & Werner, S. E. (2002). Collateral effects of behavioral parent training on families of children with developmental disabilities and behavior disorders. *Behavioral Interventions*, *17*(2), 75–83. https://doi.org/10.1002/bin.111

Ferdinand, R. F. (2008). Validity of the CBCL/YSR DSM-IV scales anxiety problems and affective problems. *Journal of Anxiety Disorders*, *22*(1), 126-134. https://doi.org/10.1016/j.janxdis.2007.01.008

Foley, S., Ronchi, L., Lecce, S., Feng, X., Chan, M. H., & Hughes, C. (2023). Cross-cultural equivalence of parental ratings of child difficulties during the pandemic: Findings from a six-site study. *International Journal of Methods in Psychiatric Research, 32*(1), e1933. https://doi.org/10.1002/mpr.1933

Forehand, R., Lafko, N., Parent, J., & Burt, K. B. (2014). Is parenting the mediator of change in behavioral parent training for externalizing problems of youth? *Clinical Psychology Review*, *34*(8), 608–619. https://doi.org/10.1016/j.cpr.2014.10.001

Ferguson, S. L., G. Moore, E. W., & Hull, D. M. (2020). Finding latent groups in observed data: A primer on latent profile analysis in Mplus for applied researchers. *International*

*Journal of Behavioral Development*, *44(5),* 458-468. https://doi.org/10.1177/01650254198817

Gach, E. J., Ip, K. I., Sameroff, A. J., & Olson, S. L. (2018). Early cumulative risk predicts externalizing behavior at age 10: The mediating role of adverse parenting. *Journal of the Division of Family Psychology, 32*(1), 92–102. https://doi.org/10.1037/FAM0000360

Goldberg, D. P., Gater, R., Sartorius, N., Ustun, T. B., Piccinelli, M., Gureje, O., & Rutter, C. (1997). The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychological Medicine*, *27*(1), 191–197. https://doi.org/10.1017/S0033291796004242

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, *38*, 581–586. https://doi.org/10.1097/00004583-200111000-00015

Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2003). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *International Review of Psychiatry*, *15*(1-2), 166-172. https://doi.org/10.1080/0954026021000046128

Goodman, S. H., Simon, H. F. M., Shamblaw, A. L., & Kim, C. Y. (2020). Parenting as a Mediator of Associations between Depression in Mothers and Children's Functioning: A Systematic Review and Meta-Analysis. *Clinical Child and Family Psychology Review*, *23*(4), 427–460. https://doi.org/10.1007/S10567-020-00322-4/METRICS

Griest, D. L., & Forehand, R. (1983). How can I get any parent training done with all these other problems going on? The role of family variables in child behavioral therapy. *Child and Family Behavior Therapy*, *4*(1), 73–80. https://doi.org/10.1300/J019v04n01_08

Groenman, A. P., Hornstra, R., Hoekstra, P. J., Steenhuis, L., Aghebati, A., Boyer, B. E., ... & van den Hoofdakker, B. J. (2022). An individual participant data meta-analysis:

behavioral treatments for children and adolescents with attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *61*(2), 144-158. https://doi.org/10.1016/j.jaac.2021.02.024

Gross, T., Alex Mason, W., Parra, G., Oats, R., Ringle, J. L., & Haggerty, K. (2015). Adherence and Dosage Contributions to Parenting Program Quality. *Journal of the Society for Social Work and Research*, *6*(4), 467–489. https://doi.org/10.1086/684108

Gutman, L. M., Joshi, H., & Schoon, I. (2019). Developmental Trajectories of Conduct Problems and Cumulative Risk from Early Childhood to Adolescence. *Journal of Youth and Adolescence*, *48*(2), 181–198. https://doi.org/10.1007/S10964-018-0971-X/FIGURES/1

Herbers, J. E., Garcia, E. B., & Obradović, J. (2017). Parenting Assessed by Observation versus Parent-report: Moderation by Parent Distress and Family Socioeconomic Status. *Journal of Child and Family Studies*, *26*(12), 3339–3350. https://doi.org/10.1007/s10826-017-0848-8

Hill, R. (1959). Generic features of families under stress. - PsycNET. *Social Casework*, *39*, 139–150. https://psycnet.apa.org/record/1959-08206-001

Holland, P. W., Dorans, N. J., & Brennan, R. L. (2006). *Educational measurement*. Praeger Publishers, Westport, CT.

Howard, M. C., & Hoffman, M. E. (2018). Variable-Centered, Person-Centered, and Person-Specific Approaches. *Organizational Research Methods*, *21*(4), 846–876. https://doi.org/10.1177/1094428117744021

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Jacobson, N. S., Roberts, L. J., Berns, S. B., & McGlinchey, J. B. (1999). Methods for defining and determining the clinical significance of treatment effects: description, application, and alternatives. *Journal of Consulting and Clinical Psychology*, *67*(3), 300–307. https://doi.org/10.1037//0022-006X.67.3.300

Kazdin, A. E., & McWhinney, E. (2018). Therapeutic alliance, perceived treatment barriers, and therapeutic change in the treatment of children with conduct problems. *Journal of Child and Family Studies*, *27*(1), 240–252. https://doi.org/10.1007/s10826-017-0869-3

Kjøbli, J., Naerde, A., Bjørnebekk, G., & Askeland, E. (2013). Maternal mental distress influences child outcomes in brief parent training. *Child and Adolescent Mental Health*, *19*(3), 171–177. https://doi.org/10.1111/camh.12028

Klahr, A. M., & Burt, S. A. (2014). Practitioner Review: Evaluation of the known behavioral heterogeneity in conduct disorder to improve its assessment and treatment. *Journal of Child Psychology and Psychiatry*, *55*(12), 1300–1310. https://doi.org/10.1111/jcpp.12268

Kreppner, K., & Lerner, M. (1989). Family systems and life-span development: Issues and perspectives. In K. Kreppner & M. Lerner (Eds.), *Family systems and life-span development* (pp. 1–13). Lawrence Erlbaum Associates Publishers.

Lane, C., Hogg, E., Karwatowska, L. A., French, L., Ranieri, V. F., Jesnick, L. G. D., Roberts, C., Scott, S., Senior, R., Skinner, G. C., & Kennedy, E. M. M. (2023). Personalised interventions for subgroups of children with conduct problems. *Cochrane Database of Systematic Reviews*, *2023*(4). https://doi.org/10.1002/14651858.CD012746.PUB2

Lasa, L., Ayuso-Mateos, J. L., Vázquez-Barquero, J. L., Dıez-Manrique, F. J., & Dowrick, C. F. (2000). The use of the Beck Depression Inventory to screen for depression in the general population: a preliminary analysis. *Journal of affective disorders*, *57*(1-3), 261-265. https://doi.org/10.1016/S0165-0327(99)00088-9

Lanza, S. T., & Rhoades, B. L. (2013). Latent Class Analysis: An Alternative Perspective on Subgroup Analysis in Prevention and Treatment. *Prevention Science*, *14*(2), 157–168. https://doi.org/10.1007/s11121-011-0201-1

Leijten, P., Gardner, F., Landau, S., Harris, V., Mann, J., Hutchings, J., Beecham, J., Bonin, E.-M., & Scott, S. (2018). Research Review: Harnessing the power of individual participant data in a meta-analysis of the benefits and harms of the Incredible Years parenting program. *Journal of Child Psychology and Psychiatry*, *59*(2), 99–109. https://doi.org/10.1111/jcpp.12781

Leijten, P., Gardner, F., Melendez-Torres, G. J., van Aar, J., Hutchings, J., Schulz, S., Knerr, W., & Overbeek, G. (2018). What to teach parents to reduce disruptive child behavior: Two meta-analyses of parenting program components. *Journal of the American Academy of Child & Adolescent Psychiatry*, *58*(2), 180–190. https://doi.org/10.1016/J.JAAC.2018.07.900

Leijten, P., Gardner, F., Melendez-Torres, G. J., van Aar, J., Hutchings, J., Schulz, S., Knerr, W., & Overbeek, G. (2019). Meta-Analyses: Key Parenting Program Components for Disruptive Child Behavior. *Journal of the American Academy of Child and Adolescent Psychiatry*, 58 (2), 180–190). https://doi.org/10.1016/j.jaac.2018.07.900

Leijten, P., Raaijmakers, M., Wijngaards, L., Matthys, W., Menting, A., Hemink-van Putten, M., & Orobio de Castro, B. (2018). Understanding Who Benefits from Parenting Interventions for Children's Conduct Problems: an Integrative Data Analysis. *Prevention Science*, *19*(4), 579–588. https://doi.org/10.1007/s11121-018-0864-y

Leijten, P., Scott, S., Landau, S., Harris, V., Mann, J., Hutchings, J., Beecham, J., & Gardner, F. (2020). Individual Participant Data Meta-analysis: Impact of Conduct Problem Severity, Comorbid Attention-Deficit/Hyperactivity Disorder and Emotional Problems, and Maternal Depression on Parenting Program Effects. *Journal of the American Academy of*

*Child & Adolescent Psychiatry*, *59*(8), 933–943. https://doi.org/10.1016/J.JAAC.2020.01.023

Liang, L. A., Berger, U., & Brand, C. (2019). Psychosocial factors associated with symptoms of depression, anxiety and stress among single mothers with young children: A population-based study. *Journal of Affective Disorders*, *242*, 255–264. https://doi.org/10.1016/J.JAD.2018.08.013

Lundin, A., Hallgren, M., & Forsell, Y. (2015). The validity of the symptom checklist depression and anxiety subscales: A general population study in Sweden. *Journal of Affective Disorders*, *183*, 247-252. https://doi.org/10.1016/j.jad.2015.05.024

Malcolm-Smith, S., Lake, M. T., Krwece, A., Du Plooy, C. P., Hoffman, N., Donald, K. A., Zar, H. J., & Stein, D. J. (2022). Externalizing behavior in preschool children in a South African birth cohort: Predictive pathways in a high-risk context. *Development and Psychopathology*, 1–18. https://doi.org/10.1017/S095457942200027X

Masarik, A. S., & Conger, R. D. (2017). Stress and child development: A review of the Family Stress Model. In *Current Opinion in Psychology* (Vol. 13, pp. 85–90). Elsevier. https://doi.org/10.1016/j.copsyc.2016.05.008

Maurice-Stam, H., Haverman, L., Splinter, A., van Oers, H. A., Schepers, S. A., & Grootenhuis, M. A. (2018). Dutch norms for the Strengths and Difficulties Questionnaire (SDQ) - parent form for children aged 2-18years. *Health and Quality of Life Outcomes*, *16*(1), 1–11. https://doi.org/10.1186/s12955-018-0948-1

McGilloway, S., Mhaille, G. N., Bywater, T., Furlong, M., Leckey, Y., Kelly, P., ... & Donnelly, M. (2012). A parenting intervention for childhood behavioral problems: a randomized controlled trial in disadvantaged community-based settings. *Journal of Consulting and Clinical Psychology*, *80*(1), 116. https://doi.org/10.1037/a0026304

McElroy, E., Villadsen, A., Patalay, P., Goodman, A., Richards, M., Northstone, K., Fearon, P., Tibber, M., Gondek, D., & Ploubidis, G.B. (2020). Harmonisation and Measurement Properties of Mental Health Measures in Six British Cohorts. London, UK: CLOSER. https://closer.ac.uk/wp-content/uploads/210715-Harmonisation-measurement-properties-mental-health-measures-british-cohorts.pdf

McKay, K., Kennedy, E., Senior, R., Scott, S., Hill, J., Doolan, M., ... & Young, B. (2020). Informing the personalisation of interventions for parents of children with conduct problems: a qualitative study. *BMC psychiatry, 20*(1), 1-12. https://doi.org/10.1186/s12888-020-02917-1

McKay, K., Kennedy, E., & Young, B. (2021). "Sometimes I think my frustration is the real issue": A qualitative study of parents' experiences of transformation after a parenting programme. *Plos one, 16*(10), e0258528. https://doi.org/10.1371/journal.pone.0258528

McMahon, R. J., Goulter, N., & Frick, P. J. (2021). Moderators of Psychosocial Intervention Response for Children and Adolescents with Conduct Problems. *Journal of Clinical Child & Adolescent Psychology*. https://doi.org/10.1080/15374416.2021.1894566

Mingebach, T., Kamp-Becker, I., Christiansen, H., & Weber, L. (2018). *Meta-meta-analysis on the effectiveness of parent-based interventions for the treatment of child externalizing behavior problems*. *13*(9), e0202855. https://doi.org/10.1371/journal.pone.0202855

Minuchin, P. (1988). Relationships within the family: A systems perspective on development. In R. A. Hinde & J. Stevenson-Hinde (Eds.), *Relationships within families: Mutual influences* (pp. 7–26). Oxford University Press.

Mullola, S., Brooks-Gunn, J., Elovainio, M., Hakulinen, C., Schneper, L. M., & Notterman, D. A. (2021). Early childhood psychosocial family risks and cumulative dopaminergic sensitizing score: Links to behavior problems in U.S. 9-year-olds. *Journal of Affective Disorders*, *280*, 432–441. https://doi.org/10.1016/j.jad.2020.11.026

Muthén, L. K., & Muthén, B. O. (2017). *Mplus User's Guide. Seventh Edition.* (7th ed.). Muthén & Muthén.

Nuevo, R., Dunn, G., Dowrick, C., Vázquez-Barquero, J. L., Casey, P., Dalgard, O. S., ... & Ayuso-Mateos, J. L. (2009). Cross-cultural equivalence of the Beck Depression Inventory: a five-country analysis from the ODIN study. *Journal of Affective Disorders, 114*(1-3), 156-162. https://doi.org/10.1016/j.jad.2008.06.021

Pelham, W. E., Dishion, T. J., Tein, J.-Y., Shaw, D. S., & Wilson, M. N. (2017). What doesn't work for whom? Exploring heterogeneity in responsiveness to the Family Check-Up in early childhood using a mixture model approach. *Prevention Science*, 1–12. https://doi.org/10.1007/s11121-017-0805-1

Pereira, M., Negrão, M., Soares, I., & Mesman, J. (2014). Decreasing harsh discipline in mothers at risk for maltreatment: A randomized control trial. *Infant Mental Health Journal*, *35*(6), 604–613. https://doi.org/10.1002/imhj.21464

Perlstein, S., Fair, M., Hong, E., & Waller, R. (2023). Treatment of childhood disruptive behavior disorders and callous-unemotional traits: a systematic review and two multilevel meta-analyses. *Journal of Child Psychology and Psychiatry*. https://doi.org/10.1111/JCPP.13774

Petersen, I. T., Bates, J. E., Dodge, K. A., Lansford, J. E., & Pettit, G. S. (2015). Describing and predicting developmental profiles of externalizing problems from childhood to adulthood. *Development and Psychopathology*, *27*(3), 791–818. https://doi.org/10.1017/S0954579414000789

Piquero, A. R., Jennings, W. G., Diamond, B., Farrington, D. P., Tremblay, R. E., Welsh, B. C., & Gonzalez, J. M. R. (2016). A meta-analysis update on the effects of early family/parent training programs on antisocial behavior and delinquency. *Journal of Experimental Criminology*, *12*(2), 229–248. https://doi.org/10.1007/s11292-016-9256-0

Reid, M. J., Webster-Stratton, C., & Baydar, N. (2004). Halting the development of conduct problems in Head Start children: The effects of parent training. *Journal of Clinical Child and Adolescent Psychology*, *33*(2), 279-291. https://doi.org/10.1207/s15374424jccp3302_10

Riglin, L., Agha, S. S., Eyre, O., Jones, R. B., Wootton, R. E., Thapar, A. K., ... & Thapar, A. (2021). Investigating the validity of the Strengths and Difficulties Questionnaire to assess ADHD in young adulthood. *Psychiatry Research*, *301*, 113984. https://doi.org/10.1016/j.psychres.2021.113984

Stevanovic, D., Jafari, P., Knez, R., Franic, T., Atilola, O., Davidovic, N., ... & Lakic, A. (2017). Can we really use available scales for child and adolescent psychopathology across cultures? A systematic review of cross-cultural measurement invariance data. *Transcultural Psychiatry*, *54*(1), 125-152. https://doi.org/10.1177/136346151668921

Rosa-Justicia, M., Saam, M. C., Flamarique, I., Borràs, R., Naaijen, J., Dietrich, A., ... & Castro-Fornieles, J. (2022). Subgrouping children and adolescents with disruptive behaviors: symptom profiles and the role of callous–unemotional traits. *European Child & Adolescent Psychiatry*, *31*, 51-66. https://doi.org/10.1007/s00787-020-01662-w

Rostad, W. L., Moreland, A. D., Valle, L. A., & Chaffin, M. J. (2018). Barriers to Participation in Parenting Programs: The Relationship between Parenting Stress, Perceived Barriers, and Program Completion. *Journal of Child and Family Studies*, *27*(4), 1264–1274. https://doi.org/10.1007/s10826-017-0963-6

Schmeck, K., Poustka, F., Döpfner, M., Plück, J., Berner, W., Lehmkuhl, G., ... & Lehmkuhl, U. (2001). Discriminant validity of the child behaviour checklist CBCL-4/18 in German samples. *European Child & Adolescent Psychiatry*, *10*, 240-247. https://doi.org/10.1007/s007870170013

Scott, S., & Dadds, M. R. (2009). Practitioner Review: When parent training doesn't work: theory-driven clinical strategies. *Journal of Child Psychology and Psychiatry*, *50*(12), 1441–1450. https://doi.org/10.1111/j.1469-7610.2009.02161.x

Scott, S., & O'Connor, T. G. (2012). An experimental test of differential susceptibility to parenting among emotionally-dysregulated children in a randomized controlled trial for oppositional behavior. *Journal of Child Psychology and Psychiatry*, *53*(11), 1184–1193. https://doi.org/111/j.1469-7610.2012.02586.x

Smith, J. D., Berkel, C., Hails, K. A., Dishion, T. J., Shaw, D. S., & Wilson, M. N. (2018). Predictors of Participation in the Family Check-Up Program: a Randomized Trial of Yearly Services from Age 2 to 10 Years. *Prevention Science*, *19*(5), 652–662. https://doi.org/10.1007/S11121-016-0679-7/METRICS

Stevanovic, D., Jafari, P., Knez, R., Franic, T., Atilola, O., Davidovic, N., ... & Lakic, A. (2017). Can we really use available scales for child and adolescent psychopathology across cultures? A systematic review of cross-cultural measurement invariance data. *Transcultural Psychiatry*, *54*(1), 125-152. https://doi.org/10.1177/1363461516689215

Tan, T. X., Gelley, C. D., & Dedrick, R. F. (2015). Non-child-related family stress, parenting styles, and behavior problems in school-age girls adopted from China. *Journal of Child and Family Studies*, *24*(10), 2881–2891. https://doi.org/10.1007/s10826-014-0092-4

Theunissen, M. H., de Wolff, M. S., & Reijneveld, S. A. (2019). The strengths and difficulties questionnaire self-report: a valid instrument for the identification of emotional and behavioral problems. *Academic Pediatrics*, *19*(4), 471-476. https://doi.org/10.1016/j.acap.2018.12.008

UNESCO Institute of Statistics. (2011). *International Standard Classification of Education 2011*. http://www.uis.unesco.org

Vugteveen, J., de Bildt, A., Theunissen, M., Reijneveld, S. A., & Timmerman, M. (2021). Validity aspects of the strengths and difficulties questionnaire (SDQ) adolescent self-report and parent-report versions among Dutch adolescents. *Assessment*, *28*(2), 601-616. https://doi.org/10.1177/1073191119858416

Weber, L., Kamp-Becker, I., Christiansen, H., & Mingebach, T. (2019). Treatment of child externalizing behavior problems: a comprehensive review and meta–meta-analysis on effects of parent-based interventions on parental characteristics. *European Child and Adolescent Psychiatry*, *28*(8), 1025–1036. https://doi.org/10.1007/S00787-018-1175-3/METRICS

Webster-Stratton, C. (2015). The Incredible Years® series: A developmental approach. In *Family-Based Prevention Programs for Children and Adolescents* (pp. 54-79). Psychology Press.Webster-Stratton, C. (2001). Parenting practices interview. *Unpublished Assessment Instrument*.

Webster-Stratton, C. (2008). The Incredible Years: Parents, Teachers, and Children Training Series. *Residential Treatment for Children & Youth*, *18*(3), 31–45. https://doi.org/10.1300/J007v18n03_04

Webster-Stratton, C. (2012). *Collaborating with Parents to Reduce Children's Behavior Problems: A Book for Therapists Using the Incredible Years Programs.* Seattle, WA. Incredible Years Inc.

Weeland, J., Chhangur, R. R., van der Giessen, D., Matthys, W., de Castro, B. O., & Overbeek, G. (2017). Intervention Effectiveness of The Incredible Years: New Insights Into Sociodemographic and Intervention-Based Moderators. *Behavior Therapy*, *48*(1), 1–18. https://doi.org/10.1016/j.beth.2016.08.002

Weeland, J., Helmerhorst, K. O. W., & Lucassen, N. (2021). Understanding differential effectiveness of behavioral parent training from a family systems perspective: Families

are greater than “ some of their parts .*Journal of Family Theory & Review*, *13*(1), 34–57. https://doi.org/10.1111/jftr.12408

Weeland, J., Leijten, P., Orobio de Castro, B., Menting, A., Overbeek, G., Raaijmakers, M., Jongerling, J., & Matthys, W. (2022). Exploring Parenting Profiles to Understand Who Benefits from the Incredible Years Parenting Program. *Prevention Science*, *24*(2), 259–270. https://doi.org/10.1007/S11121-022-01364-6

Weeland, J., van Aar, J., & Overbeek, G. (2018). Dutch Norms for the Eyberg Child Behavior Inventory: Comparisons with other Western Countries. *Journal of Psychopathology and Behavioral Assessment*, *40*(2), 224–234. https://doi.org/10.1007/s10862-017-9639-1

Weidman, A. C., Steckler, C. M., & Tracy, J. L. (2017). The jingle and jangle of emotion assessment: Imprecise measurement, casual scale usage, and conceptual fuzziness in emotion research. *Emotion*, *17*(2), 267. http://dx.doi.org/10.1037/emo0000226

Table 1.
*Characteristics of Included Studies*

| Trial | Authors (year) | Country | Design | IY program | Control condition | Recruitment strategy | *n* [a] | Child age (*M*) | % boys | Disruptive behavior problems pre-intervention (*M*)[c] | Caregiver depresssion pre-intervention (*M*)[c] | % low educated | % single caregiver | % Ethnic minority |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Larsson et al. (2009) | Norway | RCT | BASIC | Wait-list | Screening on DB via outpatient psychiatric clinics | 75 | 6.58 | 79.5 | 158.04 | 7.82 | 74.3 | 35.3 | 4.0 |
| 2 | Seabra-Santos, Gaspar et al. (2016) | Portugal | RCT | BASIC | Wait-list | Screening on DB via university clinics | 122 | 4.64 | 70.7 | 185.55 | 10.10 | 47.1 | 20.5 | -[d] |
| 3 | Leijten et al. (2017) | The Netherlands | RCT | BASIC | Wait-list | Screening on DB via outpatient psychiatric clinics and schools, open recruitment | 156 | 5.59 | 67.4 | 124.21 | 18.58 | 47.1 | 7.1 | 42.9 |
| 4 | Posthumus et al., (2012) | The Netherlands | Case-control | BASIC + ADVANCED | No active intervention[b] | Screening on DB via Office for Screening and Vaccination | 144 | 4.23 | 73.5 | 120.04 | 17.32 | 10.9 | 10.6 | -[d] |
| 5 | Scott, Sylva et al. (2010) | England | RCT | BASIC | Phone helpline | Screening on DB via schools | 109 | 5.46 | 67.2 | 118.23 | 12.42 | 55.6 | 53.6 | 32.0 |
| 6 | Scott, O'Connor et al. (2010) | England | RCT | BASIC | No active intervention | Recruitment via schools | 166 | 5.16 | 53.0 | 104.92 | 8.69 | 42.6 | 40.1 | 74.4 |

*Note.* [a] Number of families from the original study that was included in this study; [b] Only families allocated to the intervention were assessed on caregiver depression and included in this study; [c] Mean sumscore (ECBI and BDI); [d] Not reported.

Table 2.
*Results Latent Profile Analyses*

| # profiles | AIC | BIC | LRT[a] | entropy | Lowest class probability | Smallest class (%) |
|---|---|---|---|---|---|---|
| 1 | 25381.700 | 25446.786 | - | - | - | - |
| 2 | 24973.170 | 25075.448 | 0.0000 | .822 | .921 | 45.70 |
| 3 | 24808.638 | 24948.107 | 0.5657 | .875 | .907 | 4.98 |
| 4 | 24673.083 | 24849.745 | 0.0129 | .880 | .837 | 4.73 |
| 5 | 23451.095 | 23664.949 | 0.1053 | .906 | .749 | 4.52 |

*Note*. [a] Not available for multiple imputation, this test is therefore based on listwise deletion data.

Table 3.
*Latent Family Risk Profiles*

| | Child risk factors | | | | Family risk factors | | | | | | Parenting risk factors | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Emotional problems* | | ADHD symptoms* | | Caregiver depression | | Age at birth focus child* | | SES* | | Caregiver use of praise | | Caregiver use of harsh discipline | |
| | *M* 95% CI | % clinical range/ at risk[a] | *M* 95% CI | % clinical range/ at risk[a] | *M* 95% CI | % clinical range/ at risk[a] | *M* 95% CI | % < 20 years of age | *M* 95% CI | % low educ-ated | *M* 95% CI | % clinical range/ at risk[a] | *M* 95% CI | % clinical range/ at risk[a] |
| High risk (*n* = 387) | 2.98-3.53 | 37.40 | 6.05-6.60 | 51.60 | 11.87-14.92 | 30.30 | 27.04-28.41 | 12.00 | 1.10-1.39 | 85.90 | 4.78-5.15 | 15.50 | 1.65-1.93 | 8.30 |
| Low risk (*n* = 384) | 2.32-2.82 | 31.30 | 4.53-5.15 | 33.10 | 10.30-13.21 | 22.10 | 30.85-31.90 | 1.10 | 2.68-2.77 | 0.00 | 4.99-5.32 | 13.20 | 1.72-2.02 | 10.40 |

*Note.* [a] For emotional problems: a sumscore above 3 (90th percentile found in (Maurice-Stam et al., 2018); For ADHD symptoms: a sumscore above 6 (90th percentile found in Maurice-Stam et al., 2018); For caregiver depression: score of 16 or higher (Beck & Beamesderfer, 1974); For caregiver use of praise: 1 SD below mean risk group (< 3.56), see Van Aar et al., 2019); For caregiver use of harsh discipline: 1 SD above mean at risk group (> 3.53), see Van Aar et al., 2019); * Significantly different between profiles

Table 4
*Profile Exploration*

| | **Study characteristics** | | **Family characteristics** | **Disruptive child behavior** |
|---|---|---|---|---|
| **Profile** | Condition (% IY) | Minority status (%)* | Family composition (% single caregivers)* | Disruptive child behavior T1* |
| High risk | 50.6-60.7 | 41.4-54.6 | 26.9-56.8 | 135.26-144.53 |
| Low risk | 52.8-62.9 | 56.2-69.6 | 16.7-25.0 | 123.58-132.73 |

*Note.* * significantly different between profiles

Table 5
*Differential Engagement and Effectiveness*

| | **Engagement in intervention*** | | **Effectiveness on disruptive behavior [a]** | | | | | **% Clinically significant improvement T1-T2[b]** | |
|---|---|---|---|---|---|---|---|---|---|
| **Family profile** | % session attended | 95% CI | β | SE | *p* | 95% CI | Cohen's *d* | Control | IY |
| High risk | 62.18 | 56.37-68.01 | -.207 | .044 | <.001 | -.293 - -.120 | -.299 | 8.70 | 25.30 |
| Low risk | 85.59 | 83.30-87.89 | -.077 | .038 | .042 | -.151 - -.003 | -.240 | 5.70 | 16.30 |

*Note.* [a] Effect of condition on child behavior at T2; [b] Effect of condition on clinically significant change in profile 1: $\beta$ = -2.66 (95% CI=-.376 - -.157) SE = .056, $p$ < .001; in profile 2 : $\beta$ = -.158 (95% CI=-.284 - -.032), SE = .064, $p$ = .014.