Case Study 1: Evidence Based Practice Report

Theme: Interventions implemented by parents.

'How effective is a level four Triple P parenting intervention at improving child behaviour in contexts where grandparents are carers?'

Summary

This review examines the effectiveness of the level four positive parenting programme (Triple P) intervention on child behavioural outcomes in families where grandparents have caregiving responsibilities. Despite increasing recognition that grandparents often play a key role in childrearing and struggle with child behaviour, this has been largely ignored in research, perhaps because of a research bias towards more typical Western childcare arrangements (Shwalb et al., 2019). Triple P is a group-delivered positive parenting intervention delivered to carers, based on social cognitive theory with five key principles: assertive discipline, realistic expectations, parental self-care, safe and engaging environments and positive learning environments (Sanders & Mazzucchelli, 2017).

This review identified and evaluated five studies (four randomised control trials and a case study). Limitations in studies were identified, especially concerning the heterogeneity in such studies. Recommendations for future research were made. There was mixed evidence as to the effectiveness of Triple P in this context, but it was difficult, due to the variation in these studies, to identify the causes of this variation. As such, a

tenuous conclusion was made; Triple P may be effective in the context of caregiving grandparents, but it is not clear what influences such effectiveness or what the size of any effects may be.

Introduction

Child behaviour problems

Poor child externalising or disruptive behaviours (referred to in this review as poor behaviour) are external behaviours that are destructive, counter-productive, defiant or do not align with social norms; they include behaviours such as aggression, impulsiveness and disobedience (Batum & Yagmurlu, 2007; Zilanawala et al., 2019).

Outcomes of poor behaviour

Poor behaviour is associated with a variety of short and long-term negative outcomes.

Firstly, poor behaviour is associated with difficulties in the school environment. Children displaying poor behaviour at ages three, five, seven and 11 are more likely to be excluded from school and truant at the ages of 11 and 14 (Villadsen et al., 2022). Similarly, poor behaviour is a predictor of school performance, perhaps due to its detrimental impact on the teacher-student relationship (Alatupa et al., 2011) or its association with being placed in lower attaining sets or streams, even when controlling for cognitive ability and previous attainment (Hallam & Parsons, 2013). School performance itself is also a predictor of a wide array of life outcomes, such as work performance and educational level (Alatupa et al., 2011).

Moreover, poor behaviours can place strain on families and parents; they are associated with poorer parental wellbeing and higher levels of stress

(Faraji-Khiavi et al., 2021; Sourander et al., 2006). This appears to be more than just the impact of having a child with special needs; Gallagher and Hannigan (2014) found that child behaviour explained the interaction between child disability and parental mental health.

Furthermore, poor behaviour is associated with poor social outcomes. Batum and Yagmurlu (2007) suggest that emotional and behavioural regulation underlie social behaviours. This perhaps explains why poor behaviour is associated with poor social adjustment (Alatupa et al., 2011), as well as more relationship problems (Sebre et al., 2023). Additionally, children who show behavioural problems are more likely to later be victims of bullying (Campbell et al., 2019).

There are also internal impacts of poor behaviour. For instance, Caspi et al. (2003) found that child behavioural problems at age three predicted emotional problems at age 26. However, it is worth noting that poor behaviour may be associated with internalising problems (Picoito et al., 2021) and the causational and directional relationship of emotional problems and poor behaviour is complex.

In part because of these associated issues in emotional wellbeing, social outcomes, family wellbeing and school achievement, poor behaviour is associated with a variety of later negative outcomes, such as psychiatric diagnoses in adulthood; aggression (Reef et al., 2011); later poor behaviour

(Sourander et al., 2006) and increased body mass index (Anderson et al., 2010).

Educational psychologists (EPs) and poor behaviour

Because of the detrimental impacts of poor child behaviour, it is important to support children displaying poor behaviour as early as possible (Alatupa et al., 2011) and educational psychologists (EPs) are often requested to help support teachers, families and children with this (Hart, 2010). Furthermore, children who might see an EP for other needs may also need support with their behaviour. For instance, poor behaviour is associated with poorer executive functioning skills, such as inhibition and attention control (Batum & Yagmurlu, 2007). Therefore, it is important that EPs have a good understanding of how to effectively intervene with child behavioural problems.

This may be even more pertinent now as EPs support children and families with the consequences of the Covid-19 pandemic. Though research is still being published, early literature suggests that child behaviour worsened after the pandemic in and outside the home, for children with and without special educational needs (Blanden et al., 2021; Wang et al., 2021; Zhang et al., 2020)

Child behaviour interventions- parents

Poor behaviour is linked with ineffective parenting practices, lower maternal attachment, familial instability and interparental conflict (Ackerman

et al., 1999; Afandi & Ismail, 2022; Faraji-Khiavi et al., 2021; Pinquart, 2017; Sourander et al., 2006). Conversely, a positive parenting environment can also beneficially impact child behaviour; Tamura et al. (2020) found that a positive parenting environment and positive parent-child relationships buffered the impact of socioeconomic status on child behaviours and emotional wellbeing. As a consequence, previous reviews have established that parenting interventions can be effective at improving child behaviour (Jeong et al., 2021; Pedersen et al., 2019) and are a potential resource for EPs to use when supporting children displaying poor behaviour and their families.

Child behaviour interventions- grandparents

However, research has predominantly ignored the potential for parental training in the context of other carers, such as grandparents. Grandparents often play a crucial role in child-rearing, especially now as longer life expectancy and better health mean that more grandparents can take on caring roles (Cantillon et al., 2021). In the United Kingdom (UK) grandparents are also the largest kinship caring group when children are removed from their birth family (Hunt, 2018). Additionally, due to the current cost of childcare in the UK, grandparents may become an increasingly important and financially viable option of childcare, especially for those who need childcare outside regular school hours (Farquharson & Olorenshaw, 2022). Furthermore, as the UK becomes more multicultural, it is possible that grandparents may take a larger role in caring. For instance, prior to the pandemic, over one-quarter of families in the UK had some degree of

grandparent care, while this was over one-half in China (Cantillon et al., 2021). It is possible that, in a more multicultural UK, grandparents will take more caring responsibilities; however, there remains theoretical until more research is published. Perhaps because of the Western focus of research, there is limited research on grandparents as caregivers (Shwalb et al., 2019).

In addition to a potential increase in grandparent carers, grandparents may struggle more with poor behaviour; grandparents tend to experience high levels of child behavioural difficulties and consequent stress (Harnett et al., 2012; Hunt, 2018). Landry-Meyer and Newman (2004) argue that grandparents parent within ambiguous role boundaries from the child's parents and the law. Both inter-carer conflict (such as between a grandparent and parent) and ambiguous roles are associated with negative child behaviours (Landry-Meyer & Newman, 2004; Salari et al., 2014). This may be further exacerbated by trauma or difficulties that the child has experienced; custodial grandparents may be parenting their grandchildren because their child has been removed from birth parental care, which may result in more behavioural and/or emotional problems (Hunt, 2018). Moreover, grandparents may struggle more with caring at an older age (Landry-Meyer & Newman, 2004) and may be less familiar with current social approaches to parenting and available support (Hunt, 2018). Consequently, families with caregiving grandparents may need additional parenting training and support (McLaughlin et al., 2017).

Therefore, parenting interventions in the context of grandparent carers may be an appropriate intervention for EPs to recommend in order to improve child behaviour.

Triple P

Background

A potential parenting intervention that has started to be applied to the grandparent carer context is the Positive Parenting Programme or Triple P.

It is broadly accepted that parenting that is characterised by warmth, care and nurture results in an array of more positive outcomes for children than those who grow up with less positive parenting practices (Sanders et al., 2014). Triple P is an intervention that aims to increase such parenting in children from birth to 16 years old (Sanders et al., 2003).

Triple P is divided into different levels based on the need of each family (Sanders et al., 2003). Each level may also contain different types of Triple P within it (Sanders et al., 2004). Levels one to three are information sharing or brief interventions while level five is for families coping with complex problems (Sanders et al., 2003). Level four, which includes all the interventions found in this review, is for parents concerned about their child's behaviour (Sanders et al., 2003). The number of sessions in level four Triple P varies by specific intervention, but they broadly contain similar methods of support: presentations or videos with information, small group activities, individual work, homework, planning and reflection time (Sanders et al. 2014).

Theoretical basis

Triple P is primarily based on Bandura's (1986) social cognitive theory (SCT, Sanders & Mazzucchelli, 2017) which focuses on the interaction between a person (including their life experiences), the environment and their behaviour (Bandura, 1989), meaning that poor behaviour or parental stress are viewed not as a result of the child or parent alone, but as a result of interactions between people, behaviour and the environment, reducing direct blame on any individual (Thomas & Zimmer-Gembeck, 2007). Using this, parents are encouraged to recognise potential causes of behaviour and opportunities to change it (Thomas & Zimmer-Gembeck, 2007). Triple P also focuses on SCT's more within-child and within-parent factors such as self-regulatory processes, self-efficacy and the ability to perform a behaviour (Schunk, 2012; Stajkovic & Luthans, 1998).

Based on this theoretical basis, Triple P is founded on five linked core principles: *assertive discipline*; *realistic expectations*; *parental self-care;* age-appropriate *safe and engaging environments*; and *positive learning environments* through spending time with children (Sanders & Mazzucchelli, 2017).

Research base

Positive parenting approaches, such as Triple P, are broadly effective in changing parenting styles and supporting children (Leijten et al., 2019). The World Health Organisation (2010) even recognised that Triple P could

reduce child maltreatment and dysfunctional parenting behaviours as well as child aggression, disruptive behaviour and later criminal behaviour.

Nogueira et al.'s (2022) systematic review and meta-analysis also found significant effects of Triple P on child behaviour which was maintained for six months. However, they also noted a significant risk of bias within Triple P research due to conflicts of interest and the lack of double blinding.

Parent interventions for grandparents

Because of the aforementioned difficulties faced by families with grandparent carers, Triple P has started to be applied to the grandparent context to support child behaviour and (grand)parent wellbeing. However, there has not been a systematic review of the research on this topic; this is the purpose of this review

Review question

Therefore, this review's question is 'How effective is a level four Triple P parenting intervention at improving child behaviour in contexts where grandparents are carers?'

Critical Review of the Evidence Base

Literature search

A systematic literature review was conducted in December 2022 across all Ovid databases, all Proquest databases, all Web of Science databases, the Cochrane collection and PubMed using the search terms in

table one. EBSCO, Campbell Collaboration, Science Direct and JSTOR were also searched, but no results were found. See Appendix A for a full list of all databases including the sub-databases. The official Triple P website (www.triplep.net) has a list of published studies on Triple P. On this site, the 'Participants' search terms from table 1 were searched.

Table 1

Search Terms Used in Literature Searching

Intervention		Participants‡‡
“positive parenting program*” † **OR** “triple p” ‡	**AND**	grandparent* **OR** grandmother* **OR** grandfather* **OR** elder* **OR** (older) N3 relative* **OR** “mother’s mother*”†† **OR** “mother’s father*” †† **OR** “father’s mother*” †† **OR** “father’s father*” †† **OR** “second-degree relative*”†† **OR** generation*

Note. *indicates the use of a wildcard. NX indicates a proximity search.

†In search engines where wildcard inside quotation marks were not allowed, "positive parenting" N1 program* was used instead

‡In search engines where p is treated as a stop word, this was not used

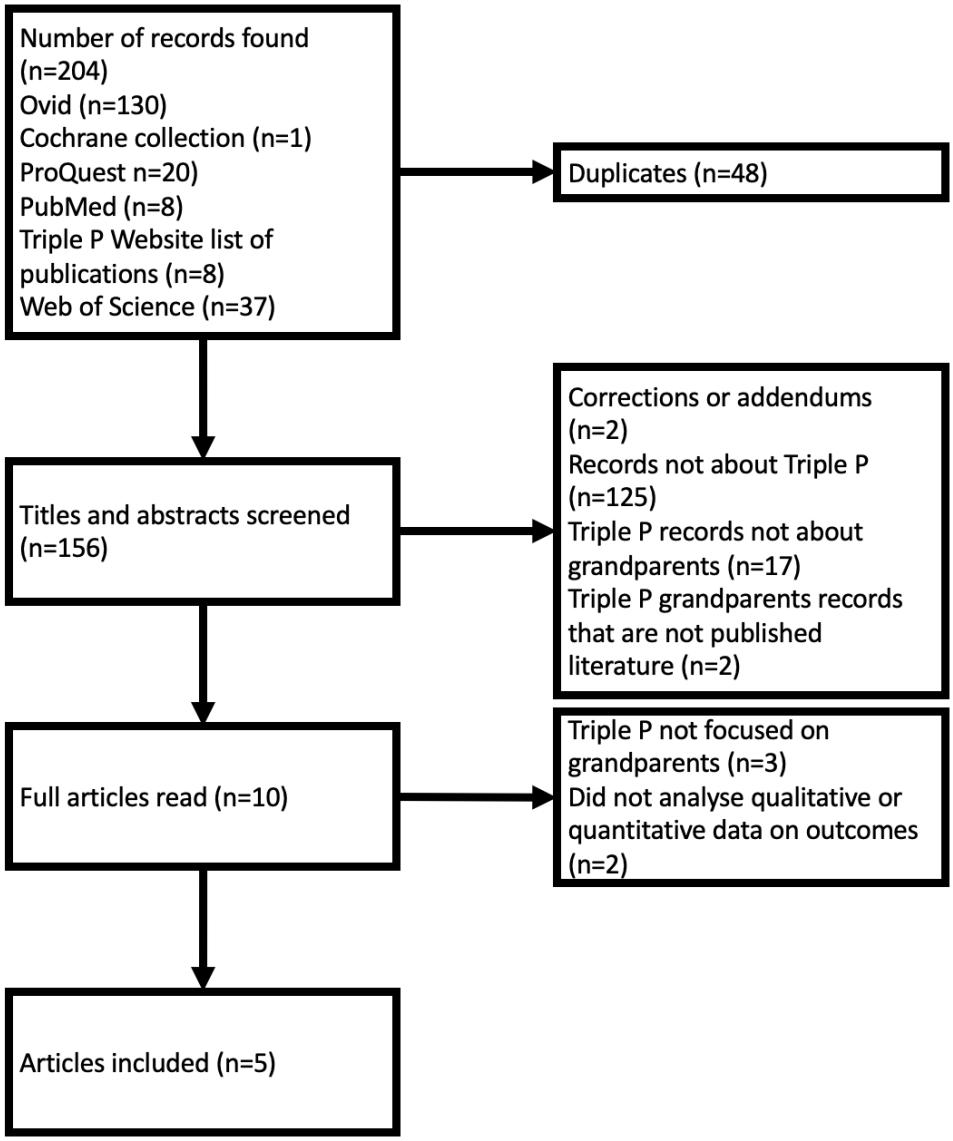
†† In search engines where wildcards inside quotation marks were not allowed, plurals were searched separately as well, e.g. “mother’s mother” OR “mother’s mothers”

‡‡ On the Triple P website, only these participant terms were used. Because this is not a search engine, these were searched using the computer’s find command function.

The search resulted in 204 results, of which 48 were duplicates (see figure one). 156 title and abstracts were screened using inclusion criteria (see table two) and 10 full articles were read (see appendix B for studies excluded at this point), resulting in five included studies (see table 3 and the appendix C Mapping the Field table for more information on each study).

Figure 1

Flowchart of the Screening Process



Note. Details of studies excluded after being read fully can be found in Appendix B.

Table 2

Inclusion and Exclusion Criteria

	Inclusion criteria	Exclusion criteria	Rationale
1. Type of publication	Peer-reviewed research article	Not peer-reviewed or not a research article	To ensure that the article has gone through a review process and quality review
2. Language	Published in English or with an English translation available	Published in a different language	To ensure that the reviewer can understand the article
3. Intervention	Intervention includes level four Triple P in context of caregiving grandparents	Does not include level four Triple P in context of caregiving grandparents	The aim is to examine Triple P focused on grandparents
4. Participants	Grandchildren whose grandparents have caring responsibilities	Is not focused on grandchildren of grandparents with caring responsibilities	The aim is to examine Triple P for grandparents
5. Outcome Measure	Child pre-post behavioural analysis is included	No measurement of child behaviour before and after the intervention	To examine the effect of the intervention on child behaviour

Note. See figure one and Appendix B for a list of studies excluded for each criteria

Table 3

Included studies

Included studies
Hoang, N. P. T., Kirby, J. N., Haslam, D. M., & Sanders, M. R. (2022). Promoting Positive Relationship Between Parents and Grandparents: A Randomized Controlled Trial of Group Triple P Plus Compassion in Vietnam. *Behavior Therapy, 53*(6), 1175-1190.
Kirby, J. N., & Sanders, M. R. (2013). Using a behavioural family intervention to produce a three-generational benefit on family outcomes: A case report. *Behaviour Change, 30*(4), 249-261.
Kirby, J. N., & Sanders, M. R. (2014). A randomized controlled trial evaluating a parenting program designed specifically for grandparents. *Behaviour Research and Therapy, 52*, 35-44.
Leung, C., Sanders, M., Fung, B., & Kirby, J. (2014). The effectiveness of the Grandparent Triple P program with Hong Kong Chinese families: A randomized controlled trial. *Journal of Family Studies, 20*(2), 104-117.
Smith, G. C., Hayslip Jr, B., Hancock, G. R., Strieder, F. H., & Montoro-Rodriguez, J. (2018). A randomized clinical trial of interventions for improving well-being in custodial grandfamilies. *Journal of Family Psychology, 32*(6), 816.

Weight of Evidence

Gough's (2007) Weight of Evidence (WoE) framework was used to assess the relevance and quality of the included studies. This framework has three components: WoE A assesses the study's methodological quality; WoE B assesses its methodological relevance; and WoE C assesses its topic relevance.

The included studies were four randomised control trials (RCTs) and a case study. For WoE A, Gersten et al.'s (2005) quality indicators were used for the RCTs while Horner et al.'s (2005) quality indicators was used for the case study (see appendix C for rationale and scoring).

WoE B was assessed using the effectiveness evidence hierarchy taken from Petticrew and Roberts' (2003) adaption of Muir Grey's (1996) evidence hierarchy (see appendix E). WoE C was assessed against specific criteria created for this review question (see appendix F). Together, these were averaged to form WoE D, the overall weight of evidence: each WoE score was given equal weighting (see table 4 or appendix G for a full breakdown of scores).

Table 4

Weight of Evidence (Gough, 2007) scores for each study

	Methodological Quality (WoE A)	Methodological Relevance (WoE B)	Topic Relevance (WoE C)	Overall Weight of Evidence (WoE D)
Hoang et al. (2022)	3 (High)	3 (High)	2.0 (Medium)	2.7 (High)
Kirby and Sanders (2013)	1.7* (Medium)	1 (Low)	2.9 (High)	1.9 (Medium)
Kirby and Sanders (2014)	1 (Low)	3 (High)	2.6 (High)	2.2 (Medium)
Leung et al. (2014)	3 (High)	3 (High)	2.4 (High)	2.8 (High)
Smith et al. (2018)	3 (High)	3 (High)	2.1 (Medium)	2.7 (High)

Note. 1-1.6 (low), 1.7-2.3 (medium), 2.4-3 (high)

*Kirby and Sanders (2013) is a case study and, therefore, Horner et al.'s (2005) rating system was used for WoE A. Gersten et al.'s (2005) indicators were used for the other studies.

All studies received high or medium WoE D, in large part because of the RCT design of four of the studies. The studies with the highest relevance to this review received the lowest WoE B and WoE A scores due to Kirby and Sanders (2013) being a case study and methodological issues in Kirby and Sanders' (2014) study. The other three studies received a high score for WoE A and WoE B but were less relevant to the study.

Interventions

The type of Triple P varied and can be divided into two categories: Triple P given to parents in the context of grandparents being carers and Triple P given to grandparents.

Only Hoang et al.'s (2022) intervention fit into the former. They delivered, to parents, a level four group Triple P approach alongside a compassion-based therapy, where parents were taught self-compassion and compassion for the caregiving grandparent. The use of the compassion therapy as well as Triple P is reflected in Hoang et al.'s (2022) low WoE C score.

Triple P was given to grandparents for the remaining studies. Smith et al. (2018) delivered a level four group intervention to grandparents, but this was not specifically tailored to their needs, resulting in a lower WoE C score. The other studies received Grandparent Triple P (GTP) which met the WoE C criteria of being solely delivered to and focused on grandparents. GTP has the same theoretical background as regular Triple P interventions but has

additional components to address the needs of grandparents (Kirby & Sanders, 2014). Kirby and Sanders (2013, 2014) created and implemented the intervention while Leung et al. (2014) adapted GTP for their Hong Kong context, including translating it into Cantonese.

Design

All studies except Kirby and Sanders (2013) followed an RCT design and were given the highest rating, three, for WoE B. Kirby and Sanders (2013) was a single-person case study and received the lowest score of one.

All studies except Leung et al. (2014) included six-month follow-up scores, reflected in their WoE C scores. The importance of including a follow-up was highlighted in Hoang et al.'s (2022) study where one outcome became non-significant while the effect size on the other increased (see table five), highlighting the interaction between time and the intervention.

All RCTs except Smith et al. (2018) used a care-as-usual or waitlist control method. Smith et al. (2018) used an information only control (IOC). They also compared Triple P to cognitive behavioural therapy though the (insignificant) results for the CBT comparison are not reported here for consistency. While the nature of the control group was not specified in WoE A, B or C, Smith et al.'s (2018) study found no significant difference between the IOC and Triple P conditions post-intervention, highlighting the potential importance of an IOC-style control.

Researchers

All of the studies except for Smith et al. (2018) were co-authored by Kirby and Sanders, both of whom work for the University of Queensland which owns Triple P. Indeed, Sanders is the founder of Triple P and, alongside Kirby, is the co-founder of GTP. This is a clear conflict of interest.

Only Smith et al. (2018) did not appear to have a conflict of interest.

Participants

Studies ranged from one to 134 participants included in the final analysis. Kirby and Sanders' (2013) case study had only one grandmother included in the analysis; it is not clear if she was also a participant in their 2014 study. Assuming she was not, across the studies, data was gathered on 345 children from 189 grandparents and 205 parents. Triple P was fully delivered to 152 grandparents and 38 parents (according to the studies' definitions of completers). Only Hoang et al. (2022) delivered Triple P to parents. Smith et al. (2018) did not provide data on completers per condition, so instead the number of participants who completed follow-up measures is used.

Children

Kirby and Sanders (2013) only included one female child. Other studies ranged from 39% to 47% female for an average of 45% female (235 male, 191 female). Boys tend to have higher levels of externalising behaviour (Midouhas, 2017) which might explain this gender gap.

All studies either specified that the children of participating (grand)parents must be below the age of twelve or the mean age of the grandchildren ± one standard deviation was between zero and 12 years. Therefore, all studies met the WoE C criteria of the children having a mean age ± one standard deviation between zero and 12 years. Kirby and Sanders (2014) and Smith et al. (2018) established age limits for their participants (two to nine years and four to 12 years, respectively). The child in Kirby and Sanders' (2013) case study was two years old. Hoang et al. (2022) and Leung et al. (2014) did not set an age limit for the participants, but the children in the Hoang et al.'s (2022) intervention were, on average, 4.54 years old (SD=2.32) and, in the control, were 4.61 years old (SD=2.52); in Leung et al.'s (2014) study, the children in the intervention were aged 3.77 years (SD=1.08) on average and in the control were aged 3.81 years (SD=1.11). See table C1 in Appendix C for the full list of means and standard deviations of children's ages.

Only one RCT met the essential criteria of WoE A that there be sufficient information to determine if the participants (e.g. the children) struggled with the difficulty presented (e.g. behaviour issues): in Hoang et al.'s (2022) study, parents had to rate their concern about their child's behaviour as a four or above on a one to five Likert scale. However, Hoang et al. (2022) introduced bias with the inclusion criteria as parents also had to indicate that they found the coparenting relationship with grandparents difficult. Therefore, only Kirby and Sanders (2013) met the criteria in WoE C

that children have a clearly identified and measured behavioural problem. This was due to the single-case design. Kirby and Sanders (2014) included grandparents who were concerned about their grandchild's behavioural problems (but did not rate this) or grandparents who were suffering from anxiety or depression. Leung et al.'s (2014) and Smith et al.'s (2018) inclusion criteria did not include any concerns about the child's behaviour. However, Smith et al. (2018) used fully custodial grandparents in the absence of birth parents: a sample who tend to experience high levels of grandchild behavioural difficulties and tend to find these behavioural difficulties challenging (Harnett et al., 2012). Participants also had to identify as White, Black, or Hispanic; it is unclear why. Leung et al.'s (2014) sample only required the child to live with their parent and/or grandparents and for both parents and grandparents to be resident in Hong Kong. This may explain why Leung et al. (2014) found notably lower pre-intervention Eyberg Child Behavior Inventory (ECBI) problem scores for the intervention group (M=11.28 (8.15)) than Kirby and Sanders' (2014) sample (M=13.28 (7.06)), perhaps explaining the lower effect sizes for Leung's study (see table 5).

(Grand)parents

Grandparent and parent participants were overwhelmingly female: Smith et al.'s (2018) and Kirby and Sanders' (2013) studies had solely female participants due to inclusion criteria and the nature of a single case study, respectively. The others varied from 80% to 94% female, aligning with gender biases observed in (grand)parenting (Horsfall & Dempsey, 2015).

The inclusion criteria of the grandparent and parent participants varied significantly.

Kirby and Sanders (2014) and Smith et al. (2018) established criteria for grandparents with multiple grandchildren: they were asked to identify the child who was the most challenging to care for. The other studies did not clarify how (grand)parent participants chose a target (grand)child.

Kirby and Sanders (2013) defined 12 hours as minimum grandparent care taking responsibilities, which Kirby and Sanders (2014), Leung et al. (2014) and Hoang et al. (2022) then followed. However, Smith et al. (2018) focused specifically on grandmothers who cared for their grandchildren without their parents for at least three months. All therefore met the 11 hours per week caring criteria from WoE C. Landry-Meyer and Newman (2004) found that legal custody for the children eased the role conflicts when caring for grandchildren, suggesting that the role of a custodial grandparent may have clearer boundaries, but because Smith et al. (2018) was the only study focused on custodial grandparents, it was not possible to analyse the effect of this; this could be an area of future research.

Context

The decision was not taken to limit studies by cultural context, given the increasing multicultural nature of the United Kingdom (UK), the historical focus on Western family structures which has omitted broader family structures from parenting research and the need for more research that is not

culturally tied (Hoang & Kirby, 2020). Furthermore, Triple P (though not yet in the context of grandparents as carers) has been found to be culturally acceptable and effective across different cultural contexts (Turner et al., 2020).

Three included studies took place in Westernised nations: Kirby and Sanders (2013, 2014) took place in Australia while Smith et al. (2018) took place in the United States of America. Hoang et al. (2022) and Leung et al. (2014) took place in South East Asia: Vietnam and Hong Kong, respectively. Hoang and Kirby (2020) suggested within Asian cultures, grandparent caregiving was an expectation; grandparents and parents filled similar roles; and grandparents and elders hold a higher level of respect than in Western cultures. It was not yet possible to determine if this affected Triple P's effectiveness, but this could be examined in the future.

Measures

There were a variety of behaviour measures used and all except those used in Smith et al.'s (2018) study are well validated and reliable. The studies on GTP (Kirby & Sanders, 2013, 2014; Leung et al., 2014) all used the ECBI problem and intensity scales which is well validated and is reliable (Abrahamse et al., 2015); Funderburk et al. (2003) found retest reliability over 10 months of r=.75 (p<.0001) and concurrent validity with the Preschool Behaviour Questionnaire - Teacher version (r=.53, p < .0001). Colvin et al. (1999) found internal consistencies Cronbach's alphas of .95 for the intensity scale and .93 for the problem scale. It was also appropriate for the contexts;

it has been translated and validated for the Hong Kong context in Leung et al.'s (2014) study (Leung et al., 2003) as well as the Australian context for Kirby and Sanders' (2013, 2014) studies (Werba, 2002).

Hoang et al. (2022) used the strengths and difficulties questionnaire's (SDQ) prosocial behaviours subscale and the Child Adjustment and Parent Self-Efficacy Scale's (CAPES) behavioural problems subscale; again, this is well validated and reliable. Stone et al.'s (2010) review suggests that the SDQ has high concurrent validity with the child behaviour checklist and r=0.65 test-retest reliability. The Vietnamese version has been validated, showing discriminant validity between clinical and non-clinical samples on the prosocial behaviours subscale ($F[1,1517]= 227.14$, $p<.0001$, $\omega^2=.13$) (Dang et al., 2017); Hoang et al. (2022) found internal consistency of the prosocial subscale at $\alpha = .77$.

Smith et al. (2018) created a latent model of externalising behaviours using the hyperactivity-inattention and conduct problems subscales from the SDQ and externalising subscale Parent Daily Report (PDR). The subscales from the SDQ have acceptable test-retest reliability, between 0.71 and 0.66 and concurrent validity with the CBCL (0.69=0.71) (Stones et al., 2010). Keil (2007) reviewed evidence on the PDR and found that it had acceptable test-retest reliability and concurrent validity with home observation and parent ratings of child behaviour. Smith et al. (2018) report that the PDR externalising subscale α was .98 and the SDQ externalising α was .75. However, Smith et al. (2018) comment that the PDR used in their study is a

modified version but do not specify how or why this was modified. Therefore, it is not possible to accurately understand the PDR's validity or reliability in this context. It is possible that Smith et al. (2018) were referring to the PDR's use with a grandparent population or that way in which the researchers took the average of an in-person and phone call for each time point. Moreover, although Smith et al. (2018) report good internal consistency for the measures included in the latent construct, there is no quantitative analysis of the reliability of validity for the latent construct itself.

The lack of reported internal consistency scores is reflected in Smith et al.'s (2018), Hoang et al. (2022) and Kirby and Sanders (2014) WoE A desirable characteristic scores (see table G1 in appendix G). Moreover, the same studies did not provide information within their studies on the measures validity, resulting in a lower score in the fourth desirable characteristic. It is important to note, however, that while Hoang et al. (2022) and Kirby and Sanders (2014) did not provide this data within their published article, the data can be found elsewhere and this validity data is outlined above. In contrast, there is no data available on the validity of Smith et al.'s (2018) modified PDR or latent construct.

Raters of children's behaviour varied: Smith et al. (2018) and Hoang et al. (2022) only used the participating adults (grandparents and parents respectively) while Kirby and Sanders (2013, 2014) used both parent and grandparents as raters. However, in Leung et al.'s (2014) RCT, despite being delivered to grandparents, only the parents rated the child's behaviour. This

inconsistency with raters limits the extent to which results can be compared as there were clearly noted difference in parent and grandparent scores. For example, in Kirby and Sanders' (2014) study at the six month follow-up, grandparents noticed a large, significant change on the ECBI problem measure while the parents did not notice a significant change (see table five). It also is possible that, when raters were also participants, they may over-estimate the effect of such interventions. Again, due to the inconsistency and lack of double raters it is difficult to see if this was the case: if it was, this would apply to Hoang et al.'s (2022) participants, the grandparents in Kirby and Sanders' (2013, 2014) studies and participants in Smith et al.'s study but not Leung et al. (2014) participants. Multiple raters, as in Kirby and Sanders' (2013, 2014) studies might mitigate such effects.

Another concern about the raters is that no research could be found which provided validity or reliability data for any measures when used with a grandparent population. Again, this is concerning given Kirby and Sanders' (2013, 2014) findings that parents and grandparents perceived a difference in outcomes. Furthermore, the interrater reliability of the measures (although none examined grandparents) suggest that there may be issues applying a standardised measure to a new population. For instance, the interrater between teachers and parents for the relevant subscales from the SDQ is only between 0.26 and 0.47. This could imply that this measure may provide different results between different raters such as grandparents and parents (Stones et al., 2010).

Such issues with raters are reflected in WoE C, where no study met the criterion E that information on children's behaviour is triangulated across settings with teacher, parent and grandparent reports.

Results

Statistical methods

There was heterogeneity in study and statistical methods. Hoang et al. (2022) and Kirby and Sanders (2014) used univariate analysis of variances (ANOVAs), examining the time x condition interaction. Hoang et al. (2022) also used multivariate ANOVAs (MANOVAs) to examine the overall impacts across variables. Leung et al. (2014) used an Analysis of Covariance (ANCOVA), with group as the independent variable, post-intervention scores as dependent variables and pre-intervention scores as covariates. Smith et al. (2018) used t-tests on change scores. Unfortunately, this heterogeneity meant that it was not possible to perform a meta-analysis.

Only Hoang et al. (2022) corrected for familywise error (using the Holm-Bonferroni correction). This is concerning given the large number of measures used by each study. Moreover, Hoang et al. (2022) were the only researchers to use a MANOVA instead of repeated ANOVAS; again, this risks the possibility of increased type one error. Corrections for familywise error was not a criterion in either Gersten et al. (2005) or Horner et al.'s (2005) quality indicators and so these issues are not reflected in WoE A scores but should be considered when examining the results in table five and

when comparing between Hoang et al.'s (2022) results and the non-corrected results from the other studies.

Effect size calculations

While statistical method differences make the studies difficult to directly compare, standardised effect sizes (Cohn's d) can be seen in table five. Hoang et al. (2022), Kirby and Sanders (2014), Leung (2014) and Smith et al. (2018) all provided effect sizes in Cohen's d. Kirby and Sanders (2013), likely because of the single case design, did not calculate a Cohen's d effect size. Using data from the study, this was calculated for this review and can be found in table five. However, researchers like Evans et al. (1998), suggest that, at the individual case level, an effect size offers less useful and easily understood data than the reliable change index (RCI) score. While reliable changes were provided in Kirby and Sanders' (2013) paper, there was not sufficient information on normative reference groups or statistics and therefore these were recalculated for this study (see table six). Werba's (2002) Australian norm data for three year olds was used because the child was aged somewhere between two and three and a half across the study (exact age is not known) and there is no data for two year olds. No grandparent normed data could be found, so parent report was used for both the grandmother and mother.

Significance and effect sizes

Post-intervention, excluding the case study, there were two non-significant changes and seven significant changes. At six-months, excluding

the case-study, the number of non-significant results increased to four, while there were three significant changes (see table five).

For the significant results, effect sizes ranges from medium to large at post-intervention and the six-month follow-up. All were positive, indicating an improvement in child behaviour. The largest effect sizes were from Kirby and Sanders (2014) study. It is unclear if grandparents or parents were more likely to report significant changes; the grandparent in Kirby and Sanders (2013) study reported fewer improvements in child behaviour than the parent while the reverse was true for Kirby and Sanders' (2014) study. Similarly, it is unclear if raters who were also participants reported higher effect sizes than raters who were not participants because only Kirby and Sanders (2013, 2014) and Leung (2014) provided data from non-participating carers. Reporters marked with an asterix in table five were also participants in Triple P.

Smith et al. (2018) was the only RCT study that did not have any significant results post-intervention. It also had the only (non-significant) negative effect size, indicating a deterioration in child behaviour. This is perhaps because Smith et al. (2018) was the only study that used an IOC rather than a waitlist control. However, Smith et al.'s (2018) results *were* significant at the six month follow up, highlighting the potential perceived or genuine impact of information and attention to improve parenting practices in the short term.

Kirby and Sanders' (2013) case study did not have any significant results according to post-hoc t-tests. However, due to the nature of this small n study, such results are only based on two people's reports and may offer a limited understanding of the effect of the study. See table six for RCI scores which may provide a better representation of the effect of the intervention.

Table 5

Effect sizes in Cohen's d (ES) at post-intervention and six month follow-up

Study name	Type of study	Overall WoE D Rating	Reporter	Measure	Post-intervention		Six month follow-up	
					ES and descriptor	Sig.	ES and descriptor	Sig.
Hoang et al. (2022)	RCT	2.7 (High)	Parent* (n=100)	SDQ-Prosocial	**d=0.50 (medium)**	**p=0.003**	**d=0.77 (medium)**	**p<0.001**
				CAPES-behavioural problems	**d=0.52 (medium)**	**p=0.012**	d=0.39 (small)	p=0.096
Kirby and Sanders (2013)*	Case study	1.9 (Medium)	Grandparent* (n=1) and parent (n=1) combined**	ECBI Problem	d=0.662 (medium)	p=0.5	d=0.684 (medium)	p=0.564
				ECBI Intensity	d=1.418 (large)	p=0.305	d=1.174 (large)	p=0.344
Kirby and Sanders (2014)	RCT	2.2 (Medium)	Grandparent* (n=54)	ECBI Problem	**d=0.73 (medium)**	**p=0.006**	**d=0.82 (large)**	**p<0.001**
			Parent (n=48)		**d=0.94 (large)**	**p=0.002**	d=0.23 (small)	p=0.354
			Grandparent* (n=54)	ECBI Intensity	**d=0.82 (large)**	**p<0.001**	d=0.36 (small)	p=0.055
			Parent (n=48)		**d=0.73 (medium)**	**p<0.001**	d=0.34(small)	p=0.109
Leung et al. (2014)	RCT	2.8 (High)	Parent (n=56)	ECBI Problem	**d=0.53 (medium)**	**p=0.020–0.009,**		
				ECBI Intensity	d=0.26 (small)	p=0.214–0.160,		

Smith et al. (2018)	RCT	2.7 (High)	Grandparent* (n=134 at post, 74 at six months)	Externalising latent construct	d=-0.32 (small)	p=0.18	**d=0.66 (medium)**	**p=0.03**

Note. Positive effect sizes show decrease in poor behaviour. Bold effect sizes and p values indicate significance at the p<.05 level.

*Indicates that raters of the child's behaviour were also participants in Triple P.

**Because of the small sample size, the parent and grandparent data was combined to form a Cohen's d using two reporters. Post-hoc paired t-tests were run to provide significance levels. Table 6 provides RCI scores for this study.

Table 6

RCI scores for the case study at post- intervention and six month follow-up

				Post-intervention		Six month follow-up	
Study name	Overall WoED Rating	Reporter	Measure	RCI and descriptor	Sig.	RCI and descriptor	Sig.
Kirby and Sanders (2013)	1.9 (Medium)	Grandparent (n=1)	ECBI Problem	0.00 (no reliable change)	p=1.000	**-6.37 (significant, reliable change)**	**p<0.001**
		Parent (n=1)		**57.33 (significant, reliable change)**	**p<0.001**	**63.70 (significant, reliable change)**	**p<0.001**
		Grandparent (n=1)	ECBI Intensity	1.37 (no reliable change)	p=0.172	0.95 (no reliable change)	p=.344
		Parent (n=1)		**4.31 (significant, reliable change)**	**p<0.001**	**3.78 (significant, reliable change)**	**p<0.001**

Note. Positive RCI scores show decrease in problem behaviour. Werba's (2002) Australian norm data for three year olds was used because the child was aged somewhere between two and three and a half across the study (exact age is not known) and there is no data for two year olds. No grandparent normed data could be found, so parent report was used for both the grandmother and mother.

The results from the case study's RCI scores are contradictory. At post-intervention and six-month follow-up, the parent reported significant and reliable changes. These were positive, indicating a decrease in problem behaviour. However, the grandmother reported three non-significant changes. She did report a significant change at the six month follow-up. However, this was a negative change, indicating deterioration in the child's behaviour. It seems likely that this discrepancy is due to the reporter and individual variation rather than the intervention and, therefore, despite the medium and high ratings of Kirby and Sander's (2013) study on WoE A and WoE C, respectively, its results are hard to generalise.

Conclusions and recommendations

This review examined the effectiveness of level four Triple P for families with grandparents in caregiving roles on behaviour outcomes.

Overall, the quality of studies was reasonably high: despite being a new and very limited area of study, there were four RCTs on the topic, of which three were rated as high on WoE A. Furthermore, the case study, although low on WoE B, was well-designed, receiving a high score on WoE A.

The research findings were mixed. Kirby and Sander's (2013) case study showed contradictory results from the two reporters and this, combined with the single participant nature of the study, meant that they could not reliably be attributed to the intervention.

At post-intervention, around three quarters of the results for the RCTs were insignificant, whereas this was around one half at the six-month follow-up. The significant findings were medium or large sized at both time points. Therefore, it does appear that Triple P in this context may have positive impacts on child behaviour. However, it is difficult to state the size of any such increases, especially since a meta-analysis was not possible.

The largest limitation of the research in this area is the heterogeneity in intervention type, design and measures. The inconsistency between studies makes comparisons difficult; there were different types of Triple P, participants, inclusion criteria, extent of grandparent caregiving responsibilities, measures, raters of child behaviour, cultural contexts, control groups and ages. Future research in the area should aim to use the same measures as previous studies to allow appropriate comparison. These measures should also be validated for the grandparent reporting group as no measures from this literature review were. Because of the number of these inconsistencies, it is difficult to attribute any differences in effects to any one difference. For example, it is not currently possible to determine why Kirby and Sanders (2014) found a significant effect with a medium effect size on ECBI post-intervention parent-measured intensity while Leung et al. (2014) recorded a non-significant effect on the same measure at the same time point with the same raters.

Another concern is the potential for bias given that four of the five studies involved the founders of GTP: Kirby and Sanders. Indeed, Kirby and Sanders' (2014) study had the highest effect sizes of any study. While it is not possible to conclude that this is due to bias, it is certainly concerning, especially as Nogueira et al. (2022) found similar issues of conflicts of interest when broadly reviewing Triple P. Independent research is clearly needed.

In the future, it would also be beneficial to examine the data from Triple P studies that included grandparents or families with grandparents as caregivers but did not provide this data separately (e.g. Keown et al., 2018 and Sumargi et al., 2015). If available, secondary analyses of such studies may provide additional data quickly and with limited cost.

There are implications for EP practice from this review. Firstly, EPs, when supporting families with caregiving grandparents, could consider Triple P as an intervention. Similarly, EPs working on a systemic level (for instance, working in a local authority) could consider commissioning Triple P as an intervention in this context. However, if implementing Triple P as an intervention in this context, due to the aforementioned issues in literature, it would be important to examine child outcomes, whether this is in published literature or more informal progress monitoring. Furthermore, EPs may want to consider the parenting needs of families where grandparents are carers as this review suggests that there may be a positive impact for parenting interventions in this context. Moreover, the studies delivered to grandparents

(all except Hoang et al., 2022), suggest that EPs could consider supporting caregiving grandparents, even if they are not the primary caregiver (as in Smith et al.'s 2018 study) in order to support child behaviour.

More broadly, these studies highlight the importance of the broader system around a child. While there are numerous studies on interventions for parents or schools to improve child behaviour, this review highlights that more periphery members of a child's life may have significant impacts on their behaviour. For instance, most of the participants in Kirby and Sanders' (2014) study provided less than 20 hours of childcare a week to their grandchildren, roughly equivalent to just three days a week in school. However, there was still a significant impact of a parenting intervention on these children's behaviour. In EP practice, therefore, this review highlights the importance of adults in a child's life beyond just their teachers and parents.

Overall, there are indications that Triple P can be effective for families where grandparents have caregiving roles. However, more consistent and independent research is needed before the extent of such effectiveness can be determined.

References and Appendices

References

Abrahamse, M.E., Junger, M., Leijten, P.H.O., Lindeboom, R., Boer, F. & Lindauer, 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*, 679–691 (2015). https://doi.org/10.1007/s10862-015-9482-1

Ackerman, B. P., Kogos, J., Youngstrom, E., Schoff, K., & Izard, C. (1999). Family instability and the problem behaviors of children from economically disadvantaged families. *Developmental psychology, 35(*1), 258–268. https://doi.org/10.1037/0012-1649.35.1.258

Afandi, A. A. H., & Ismail, N. (2022). Maternal Attachment, Peer Relationship and Behavior Problems among Primary School Children. *International Journal of Academic Research in Business and Social Sciences, 12*(10), 98 – 109. http://dx.doi.org/10.6007/IJARBSS/v12-i10/15198

Alatupa, S., Pulkki-Råback, L., Hintsanen, M., Mullola, S., Lipsanen, J., & Keltikangas-Järvinen, L. (2011). Childhood disruptive behaviour and school performance across comprehensive school: a prospective cohort study. *Psychology, 2*(6), 542-551. https://doi.org/10.4236/psych.2011.26084

Anderson, S. E., He, X., Schoppe-Sullivan, S., & Must, A. (2010). Externalizing behavior in early childhood and body mass index from age 2 to 12 years: longitudinal analyses of a prospective cohort study. *BMC Pediatrics*, *10*(49), 1-8. https://doi.org/10.1186/1471-2431-10-49

Bandura, A. (1986). *Social foundations of thought and action.* Englewood Cliffs.

Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist*, *44*(9), 1175-1184. https://doi.org/10.1037/0003-066X.44.9.1175

Batum, P., & Yagmurlu, B. (2007). What counts in externalizing behaviors? The contributions of emotion and behavior regulation. *Current Psychology: Developmental• Learning• Personality• Social, 25*, 272-294. https://doi.org/10.1007/BF02915236

Blanden, J., Crawford, C., Fumagalli, L., & Rabe, B. (2021). *School closures and children's emotional and behavioural difficulties.* Institute for Social and Economic Research, University of Essex. https://conference.iza.org/conference_files/covid_2021/rabe_b21456.pdf

Bradbury, A., Hoskins, K., & Fogarty, L. (2021). Policy actors in a hostile environment: the views of staff in maintained nursery schools in England. *Education 3-13, 51*(1), 171-185. https://doi.org/10.1080/03004279.2021.1956561

Campbell, M., Straatmann, V. S., Lai, E. T., Potier, J., Pinto Pereira, S. M., Wickham, S. L., & Taylor-Robinson, D. C. (2019). Understanding social inequalities in children being bullied: UK Millennium Cohort Study findings. *PloS one, 14*(5), e0217162. https://doi.org/10.1371/journal.pone.0217162

Cantillon, S., Moore, E., & Teasdale, N. (2021). COVID-19 and the Pivotal role of Grandparents: Childcare and income Support in the UK and South Africa. *Feminist Economics*, *27*(1-2), 188-202. https://doi.org/10.1080/13545701.2020.1860246

Caspi, A., Harrington, H., Milne, B., Amell, J. W., Theodore, R. F., & Moffitt, T. E. (2003). Children's behavioral styles at age 3 are linked to their adult

personality traits at age 26. *Journal of personality, 71*(4), 495-514. https://doi.org/10.1111/1467-6494.7104001

Colvin, A., Eyberg, S. M., & Adams, C. D. (1999). *Restandardization of the Eyberg Child Behavior Inventory*. University of Florida, Child Study Laboratory. https://www.researchgate.net/profile/Sheila-Eyberg/publication/239922241_Restandardization_of_the_Eyberg_Child_Behavior_Inventory/links/5429faca0cf277d58e870391/Restandardization-of-the-Eyberg-Child-Behavior-Inventory.pdf

Dang, H. M., Nguyen, H., & Weiss, B. (2017). Incremental validity of the child behavior checklist (CBCL) and the strengths and difficulties questionnaire (SDQ) in Vietnam. *Asian Journal of Psychiatry, 29*, 96-100. http://dx.doi.org/10.1016/j.ajp.2017.04.023

Dudley, J., Capous-Desyllas, M., Cisneros, T., Perez, N., & Boyns, D. (2023). Understanding the Strengths and Challenges of Grandparent Caregivers Raising Children and Youth: Creating Community through Informal Culture. *Journal of Evidence-Based Social Work*, 1-31. https://doi.org/10.1080/26408066.2022.2159778

Evans, C., Margison, F., & Barkham, M. (1998). The contribution of reliable and clinically significant change methods to evidence-based mental health. *BMJ Mental Health, 1*(3), 70-72. http://dx.doi.org/10.1136/ebmh.1.3.70

Faraji-Khiavi, F., Zahiri, M., Amiri, E., Dindamal, B., & Pirani, N. (2021). The experiences of families raising autistic children: A phenomenological study. *Journal of Education and Health Promotion, 10*(78). doi:10.4103/jehp.jehp_837_20

Farquharson, C. & Olorenshaw, H. (2022). *The changing cost of childcare*, Institute of Fiscal Studies (No.R210). https://doi.org/10.1920/re.ifs.2022.0210

Funderburk, B. W., Eyberg, S. M., Rich, B. A., & Behar, L. (2003). Further psychometric evaluation of the Eyberg and Behar rating scales for parents and teachers of preschoolers. *Early Education and Development, 14*(1), 67–81. https://doi.org/10.1207/s15566935eed1401_5

Gallagher, S., & Hannigan, A. (2014). Depression and chronic health conditions in parents of children with and without developmental disabilities: The growing up in Ireland cohort study. *Research in developmental disabilities, 35*(2), 448-454. https://doi.org/10.1016/j.ridd.2013.11.029

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

Gough, D. (2007). Weight of Evidence: A Framework for the Appraisal of the Quality and Relevance of Evidence. *Research Papers in Education, 22*(2), 213-228. https://doi.org/10.1080/02671520701296189

Hallam, S., & Parsons, S. (2013). Prevalence of streaming in UK primary schools: Evidence from the Millennium Cohort Study. *British Educational Research Journal, 39*(3), 514-544. https://doi.org/10.1080/01411926.2012.659721

Hart, R. (2010). Classroom behaviour management: Educational psychologists' views on effective practice. *Emotional and behavioural difficulties, 15*(4), 353-371. https://doi.org/10.1080/13632752.2010.523257

Hoang, N. P. T., & Kirby, J. N. (2020). A meta-ethnography synthesis of joint care practices between parents and grandparents from Asian cultural

backgrounds: Benefits and challenges. *Journal of Child and Family Studies, 29*, 605-619. https://doi.org/10.1007/s10826-019-01553-y

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional children*, *71*(2), 165-179. https://doi.org/10.1177/001440290507100203

Horsfall, B., & Dempsey, D. (2015). Grandparents doing gender: Experiences of grandmothers and grandfathers caring for grandchildren in Australia. *Journal of Sociology*, *51*(4), 1070-1084. doi: 10.1177/1440783313498945

Hunt, J. (2018). Grandparents as substitute parents in the UK. *Contemporary Social Science*, *13*(2), 175-186. https://doi.org/10.1080/21582041.2017.1417629

Jeong, J., Franchett, E. E., Ramos de Oliveira, C. V., Rehmani, K., & Yousafzai, A. K. (2021). Parenting interventions to promote early child development in the first three years of life: A global systematic review and meta-analysis. *PLoS medicine, 18*(5), e1003602. https://doi.org/10.1371/journal.pmed.1003602

Keil, V. (2007*). Psychometric properties of the parent daily report and its potential for use in child welfare settings (Doctoral thesis)* https://escholarship.org/content/qt0xh3m4q1/qt0xh3m4q1.pdf

Keown, L. J., Sanders, M. R., Franke, N., & Shepherd, M. (2018). Te Whānau Pou Toru: A randomized controlled trial (RCT) of a culturally adapted low-intensity variant of the Triple P-Positive Parenting Program for indigenous Māori families in New Zealand. *Prevention Science*, *19*, 954-965. https://doi.org/10.1007/s11121-018-0886-5

Landry-Meyer, L., & Newman, B. M. (2004). An exploration of the grandparent caregiver role. *Journal of family issues*, *25*(8), 1005-1025. https://doi.org/10.1177/0192513X04265955

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 & Adolescent Psychiatry*, *58*(2), 180-190. https://doi.org/10.1016/j.jaac.2018.07.900

Leung, C. M., Chan, S. C. M., Pang, R. C. Y., & Cheng, W. K. C. (2003). *Validation of the Chinese version of the Eyberg Child Behaviour Inventory for use in Hong Kong.* Hong Kong, China: Education Bureau, Hong Kong SAR Government.

McLaughlin, B., Ryder, D., & Taylor, M. F. (2017). Effectiveness of interventions for grandparent caregivers: A systematic review*. Marriage & Family Review, 53*(6), 509-531. https://doi.org/10.1080/01494929.2016.1177631

Midouhas, E. (2017). School poverty effects on trajectories of child behaviour: Do they depend on gender and ethnicity? *Health & place, 46*, 281-292. https://doi.org/10.1016/j.healthplace.2017.06.009

Muir Gray, J.M. (1996) *Evidence-based healthcare.* Churchill Livingstone.

Pedersen, G. A., Smallegange, E., Coetzee, A., Hartog, K., Turner, J., Jordans, M. J., & Brown, F. L. (2019). A systematic review of the evidence for family and parenting interventions in low-and middle-income countries: child and youth mental health outcomes. *Journal of Child and Family Studies, 28*, 2036-2055. https://doi.org/10.1007/s10826-019-01399-4

Petticrew, M., & Roberts, H. (2003). Evidence, hierarchies, and typologies: horses for courses. *Journal of Epidemiology & Community Health*, *57*(7), 527-529. http://dx.doi.org/10.1136/jech.57.7.527

Picoito, J., Santos, C., & Nunes, C. (2021). Emotional and behavioural pathways to adolescent substance use and antisocial behaviour: results from the UK Millennium Cohort Study. *European Child & Adolescent Psychiatry, 30*(11), 1813-1823. https://doi.org/10.1007/s00787-020-01661-x

Pinquart, M. (2017). Associations of parenting dimensions and styles with externalizing problems of children and adolescents: An updated meta-analysis. *Developmental psychology, 53*(5), 873. https://doi.org/10.1037/dev0000295

Reef, J., Diamantopoulou, S., van Meurs, I., Verhulst, F. C., & van der Ende, J. (2011). Developmental trajectories of child to adolescent externalizing behavior and adult DSM-IV disorder: results of a 24-year longitudinal study. *Social Psychiatry and Psychiatric Epidemiology*, *46*, 1233-1241. https://doi.org/10.1007/s00127-010-0297-9

Sebre, S. B., Bite, I., Miltuze, A., & Kolesovs, A. (2022). Children's relationship problems and the role of adaptive functioning, emotion dysregulation and parental cognitive restructuring. *European Journal of Developmental Psychology, 20*(2), 251-267. https://doi.org/10.1080/17405629.2022.2084067

Sanders, M. R., & Mazzucchelli, T. G. (Eds.). (2017). *The power of positive parenting: Transforming the lives of children, parents, and communities using the Triple P system*. Oxford University Press.

Sanders, M. R., Cann, W., & Markie-Dadds, C. (2003). The Triple P-Positive Parenting Programme: a universal population-level approach to the

prevention of child abuse. *Child Abuse Review: Journal of the British Association for the Study and Prevention of Child Abuse and Neglect*, *12*(3), 155-171. https://doi.org/10.1002/car.798

Sanders, M. R., Kirby, J. N., Tellegen, C. L., & Day, J. J. (2014). The Triple P-Positive Parenting Program: A systematic review and meta-analysis of a multi-level system of parenting support. *Clinical Psychology Review*, *34*(4), 337-357. https://doi.org/10.1016/j.cpr.2014.04.003

Sanders, M. R., Mazzucchelli, T. G., & Studman, L. J. (2004). Stepping Stones Triple P: The theoretical basis and development of an evidence-based positive parenting program for families with a child who has a disability. *Journal of Intellectual and Developmental Disability*, *29*(3), 265-283. https://doi.org/10.1080/13668250412331285127

Schunk, D. H. (2012). Social cognitive theory. In K. R. Harris, S. Graham, T. Urdan, C. B. McCormick, G. M. Sinatra, & J. Sweller (Eds.), *APA educational psychology handbook, Vol. 1. Theories, constructs, and critical issues* (pp. 101–123). American Psychological Association.

Shwalb, D. W., Hossain, Z., & Eisberg, G. (2019). The roles of grandparents in child development: a cultural approach. In Tulvise, T., Best, D. L., & Gibbons, J. L. *Children's social worlds in cultural context* (pp. 161-172). Springer.

Sourander, A., Pihlakoski, L., Aromaa, M., Rautava, P., Helenius, H., & Sillanpää, M. (2006). Early predictors of parent-and self-reported perceived global psychological difficulties among adolescents: a prospective cohort study from age 3 to age 15. *Social Psychiatry and Psychiatric Epidemiology, 41,* 173-182. https://doi.org/10.1007/s00127-005-0013-3

Stajkovic, A. D., & Luthans, F. (1998). Social cognitive theory and self-efficacy: Going beyond traditional motivational and behavioral approaches. *Organizational dynamics, 26*(4), 62-74. https://doi.org/10.1016/S0090-2616(98)90006-7

Stone, L. L., Otten, R., Engels, R. C., Vermulst, A. A., & Janssens, J. M. (2010). Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4-to 12-year-olds: a review. *Clinical child and family psychology review, 13,* 254-274. https://doi.org/10.1007/s10567-010-0071-2

Sumargi, A., Sofronoff, K., & Morawska, A. (2015). A randomized-controlled trial of the Triple P-Positive Parenting Program seminar series with Indonesian parents*. Child Psychiatry & Human Development, 46*, 749-761. https://doi.org/10.1007/s10578-014-0517-8

Tamura, K., Morrison, J., & Pikhart, H. (2020). Children's behavioural problems and its associations with socioeconomic position and early parenting environment: findings from the UK Millennium Cohort Study. *Epidemiology and psychiatric sciences, 29*, e155. https://doi.org/10.1017/S2045796020000700

Turner, K. M., Singhal, M., McIlduff, C., Singh, S., & Sanders, M. R. (2020). Evidence-based parenting support across cultures: the Triple P—Positive Parenting Program experience. In Halford, W., K., & Van De Vijver, F. *Cross-cultural family research and practice* (pp. 603-644). Academic Press.

Villadsen, A., Cameron, C. L., Evans, J., Van Herwegen, J., Hill, V., Hurry, J., Roberts A, Wyse D, & Johansen, T. (2022). Examination of bidirectional association of conduct and emotional problems with school exclusion and

truancy: Evidence from the UK Millennium Cohort Study, *PsyArXiv*. 10.31234/osf.io/sf3b2

Wang, L., Zhang, Y., Chen, L., Wang, J., Jia, F., Li, F., Froehlich, T. E., Hou, Y., Hao, Y., Shi, Y., Deng, H, Zhang, J., Huang, L., Xie, X., Fang, S., Xu, L., Xu, Q., Guan, H., Wang, W., Shen, J., Qian, Y., Wang, X., Shan, L., Tan, C., Yu, Y., Wang, X., Chen, F., Zhang, L., Shi, X., Li, X., & Li, T. (2021). Psychosocial and behavioral problems of children and adolescents in the early stage of reopening schools after the COVID-19 pandemic: a national cross-sectional study in China. *Translational psychiatry, 11*(1), 342. https://doi.org/10.1038/s41398-021-01462-z

Werba, B. E. (2002). *Standardization and cross-cultural validity of the Eyberg Child Behavior Inventory with Australian preschoolers (*Doctoral thesis). https://www.proquest.com/docview/276259410?pq-origsite=gscholar&fromopenview=true

World Health Organization. (2010). *Violence prevention: the evidence. https://apps.who.int/iris/bitstream/handle/10665/77936/9789241500845_eng.pdf*

Zhang, J., Shuai, L., Yu, H., Wang, Z., Qiu, M., Lu, L., Cao, X., Xia, W., Wang, Y. & Chen, R. (2020). Acute stress, behavioural symptoms and mood states among school-age children with attention-deficit/hyperactive disorder during the COVID-19 outbreak. Asian Journal of Psychiatry, 51, 102077. https://doi.org/10.1016/j.ajp.2020.102077

Zilanawala, A., Sacker, A., & Kelly, Y. (2019). Internalising and externalising behaviour profiles across childhood: The consequences of changes in the

family environment. *Social Science & Medicine, 226,* 207-216. https://doi.org/10.1016/j.socscimed.2019.02.048

Appendix A- Full list of databases searched

Table A1

List of databases and sub-databases searched

Databases	Sub databases (if relevant)
Campbell Collaboration*	
Cochrane Collection	
EBSCO*	British Education Index
	Child Development and Adolescent Studies
	Education Abstracts
	Educational Administration Abstracts
	ERIC
	Humanitities International Index
	Library and Information Science Source
	Library, Information Science and Technology Abstracts
	OpenDissertations
	Teacher Reference Center
JSTOR*	
Ovid	Allied and complementary medicine
	Books@Ovid
	Embase Classic
	Embase
	Emcare
	GEOBASE

	GEORef
	GEORef's InProcess
	Global Helath
	Health and psychosocial instruments
	ICONDA
	International pharmaceutical abstracts
	Maternity and Infant Care Database
	Ovid MEDLINE
	APA PsychArticles
	APA PsycBooks
	PsycExtra
	PsycInfo
	PsycTests
	PsycTherapy
	Social Policy and Practice
	Transplant Library
	HMIC Health Management Information Consortium
Proquest databases	Academic Video Online
	Acta Sanctorum
	American Periodicals
	Art and Architecture Archive
	The Artforum Archive
	Arts Premium Collection
	Music and Performing Arts Collection
	Screen Studies Collection

Australian Education Index

Avery Index to Architectural Periodicals

British Periodicals

The Cecil Papers

Colonial Legacies: Empire & Commonwealth Periodicals

Colonoial State Papers

Coronavirus Research Database

Digital National Security Archive

Documents on British Policy Overseas

Early Modern Books

Ebook Central

Economist Intelligence Unit Country Reports Archive

Education Magazine Archive

Entertainment Industry Magazine Archive

Gerritsen Women's History Collectoion of Aletta H Jacobs

Humanities Index

LGBT Magazine Archive

News, Policy & Politics Magazine Archive

Patrologia Latina

Periodicals Archive Online

Philosopher's Index

PRISMA Database

	ProQuest Central
	ProQuest Civil War Era
	ProQuest Dissertations and Theses Global
	ProQuest Historical Newspapers Collection
	ProQuest One Literature
	PTSDpubs
	Social Science Premium Collection
	The Vogue Archive
	Women's Magazine Archive
	The Women's Wear Daily Archive
	Youth and Popular Culture Magazine Archive
PubMed	
Science Direct*	
Triple P Website (www.triplep.net)**	
Web of Science	Web of Science Core Collection
	BIOSIS Citation Index
	Current Contents Connect
	Data Citation Index
	Derwent Innovations Index
	KCI-Korean Journal Database
	MEDLINE
	Preprint Citation Index
	SciELO Citaiton Index
	Zoological Record

*No results were found

** This is the official Triple P website (www.triplep.net) which has a list of published studies on Triple P. On this site, the 'Participants' search terms from table 1 were searched.

Appendix B- Excluded studies after full reading

Table B1

Studies excluded after full text reading

Study	Reason for exclusion
Matsumoto, Y., Sofronoff, K., & Sanders, M. R. (2010). Investigation of the effectiveness and social validity of the Triple P Positive Parenting Program in Japanese society. *Journal of Family Psychology, 24*(1), 87.	Triple P not focused on or delivered to grandparents
Cohen, B. E., Ateah, C. A., Chartier, M. J., DeCoteau, M. A., Harris, E., & Serwonka, K. (2016). Report of an equity-focused health impact assessment of a proposed universal parenting program in Manitoba. *Canadian Journal of Public Health, 107(*1), 112-118.	
Keown, L. J., Sanders, M. R., Franke, N., & Shepherd, M. (2018). Te Whānau Pou Toru: A randomized controlled trial (RCT) of a culturally adapted low-intensity variant of the Triple P-Positive	Triple P delivered to some grandparents within the sample, but their data was not given separately and the intervention was not focused on grandparents.

Parenting Program for indigenous Māori families in New Zealand. *Prevention Science, 19(*7), 954-965. Kirby, J. N., & Sanders, M. R. (2014). The acceptability of parenting strategies for grandparents providing care to their grandchildren. *Prevention Science, 15*(5), 777-787. Smith, G. C., Hancock, G. R., & Hayslip, B. (2022). Predictors and moderators of treatment efficacy in reducing custodial grandmothers' psychological distress. *Aging & Mental Health, 26*(2), 250-262.	Did not analyse pre and post qualitative or quantitative data on child behavioural outcomes

Appendix C- Mapping the Field Tables

Table C1

Mapping the Field Table- Study Design

				Intervention			Control	
Authors	Country	Design	Follow up	Type	Intervention delivered to (n)	Child mean age (SD)	Type (n)	Child mean age (SD)
Hoang et al. (2022)	Vietnam	RCT	Post and six months	Triple P plus a compassion module (Building Coparenting Alliance)	Parents (50)	4.54 (2.32)	Waitlist control (50)	4.61 (2.52)
Kirby & Sanders (2013)	Australia	Case study	Post and six months	Grandparent Triple P	Grandparent (1)	2 (N/A)	Not applicable	
Kirby & Sanders (2014)	Australia	RCT	Post and six months	Grandparent Triple P	Grandparents (28)	4.88 (2.80)	Care as usual (26)	3.92 (1.57)
Leung et al. (2014)	Hong Kong	RCT		Grandparent Triple P	Grandparents (29)	3.77 (1.08)	Waitlist control (27)	3.81 (1.11)
Smith et al. (2018)	United States of America	RCT	Post and six months	Triple P	Grandparents (115)	7.60 (2.62)	Cognitive Behavioural Therapy (CBT, 128) Information only control (IOC, 100)	CBT- 7.69 (2.61) IOC- 8.20 (2.43)

Table C2

Mapping the Field Table- Study Outcomes

Authors	Behavioural measure(s)	Measures given to (n)	Main behavioural findings*
Hoang et al. (2022)	Strengths and Difficulties Questionnaire (SDQ)- Prosocial behaviours subscale Child Adjustment and Parental Self-Efficacy Scale (CAPES) – Behavioural problems subscales	Parents (100)	Significant and medium effects (behavioural improvements) on outcomes post-intervention and at the six month follow up, except for the CAPES measure at six-months follow-up (non-significant, small effect)
Kirby & Sanders (2013)	The Eyberg Child Behavior Inventory (ECBI)	Grandparent (1) and parent (1)	Using the reliable change index, parent reported significant reliable decreases in problem behaviour at both time points. Grandparent reported no significant reliable changes, except a significant deterioration at six months.
Kirby & Sanders (2014)	The Eyberg Child Behavior Inventory (ECBI)	Grandparents (54) Parents (48)	Significant medium to large improvements post-intervention. Grandparents on the ECBI problem measure reported a significant, large improvement in behaviour at six months follow up; all other measures were insignificant and small at six months follow up.
Leung et al. (2014)	The Eyberg Child Behavior Inventory (ECBI)	Parents (56)	Significant, medium improvement on ECBI problem, but a small, non-significant change on ECBI intensity.
Smith et al. (2018)	Externalising latent construct	Grandparents (134 at post, 74 at six months)	Compared to the IOC, a small, insignificant deterioration post-intervention but a medium, significant improvement at the six month follow up.

*Significance at the p<.05 level. See table five for full significance levels and effect sizes

Appendix D- WoE A quality indicators

Gersten et al.'s (2005) quality of evidence indicators were used for the randomised controlled trials. This approach was chosen as it is appropriate for children with special educational needs (SEN) and therefore for grandchildren with behavioural issues. It has 10 essential characteristics and eight desirable characteristics (see table 2).

Gersten et al. (2005) recommends that studies may be considered high quality if they have at least nine of the 10 essential characteristics and four of the eight desirable characteristics; they can be considered acceptable quality if they have at least nine of the 10 essential characteristics and at least one of the eight desirable characteristics. Studies who met Gersten et al.'s high quality criteria were designated as '3' on the WoE while those who met the acceptable quality criteria were designated as '2' on the WoE. Studies that met more than five (half) but less than nine essential criteria were designated as '1' and studies with fewer than half of the of the essential criteria were designated as '0'. Appendix F has the full rating scores for all studies.

Table D1

Gersten et al.'s quality of evidence indicators

Essential or desirable quality indicator	Type of indicator	Criteria
Essential	Quality Indicators for Describing Participants	E1. Was sufficient information provided to determine/confirm whether the participants demonstrated the disability(ies) or difficulties presented? E2. Were appropriate procedures used to increase the likelihood that relevant characteristics of participants in the sample were comparable across conditions? E3. Was sufficient information given characterizing the interventionists or teachers provided? Did it indicate whether they were comparable across conditions?
	Quality Indicators for Implementation of the Intervention and Description of Comparison Conditions	E4. Was the intervention clearly described and specified? E5. Was the fidelity of implementation described and assessed? E6. Was the nature of services provided in comparison conditions described?
	Quality Indicators for Outcome Measures	E7. Were multiple measures used to provide an appropriate balance between measures closely aligned with the interventions and measures of generalized performance? E8. Were outcomes for capturing the intervention's effect measured at the appropriate times?
	Quality Indicators for Data Analysis	E9. Were the data analysis techniques appropriately linked to key research questions and hypotheses? Were they appropriately linked to the unit of analysis in the study? E10. Did the research report include not only inferential statistics but also effect size calculations?
Desirable		D1. Was data available on attrition rates among intervention samples? Was severe overall attrition documented? If so, is attrition comparable across samples? Is overall attrition less than 30%?

	D2. Did the study provide not only internal consistency reliability but also test–retest reliability and interrater reliability (when appropriate) for outcome measures? Were data collectors and/or scorers blind to study conditions and equally (un)familiar to examinees across study conditions? D3. Were outcomes for capturing the intervention's effect measured beyond an immediate posttest? D4. Was evidence of the criterion-related validity and construct validity of the measures provided? D5. Did the research team assess not only surface features of fidelity implementation (e.g., number of minutes allocated to the intervention or teacher/interventionist following procedures specified), but also examine quality of implementation? D6. Was any documentation of the nature of instruction or series provided in comparison conditions? D7. Did the research report include actual audio or videotape excerpts that capture the nature of the intervention? D8. Were results presented in a clear, coherent fashion?

Horner et al.'s (2005) quality indicators were used for the case study. This was chosen because of its focus on small n studies, children and special educational needs, making it appropriate for this study. For each category, the percentage of criteria met was calculated. If less than 25% of criteria were met, the study was given a score of zero for this category. If 25% or more were met but less than 50%, the study was given a score of one for this category. If 50% or more were met but less than 75%, the study was given a score of two for this category. If more than 75% were met, the study was given a score of three for this category. Each category was then averaged to create the overall score for each study. Full scores can be found in appendix F.

Table D2

Horner et al.'s (2005) quality of evidence indicators

Category	Criteria
Description of Participants and Settings	1 Participants are described with sufficient detail to allow others to select individuals with similar characteristics (e.g., age, gender, disability, diagnosis). 2 The process for selecting participants is described with replicable precision. 3 Critical features of the physical setting are described with sufficient precision to allow replication.
Dependent Variable	4 Dependent variables are described with operational precision 5 Each dependent variable is measured with a procedure that generates a quantifiable index. 6 Measurement of the dependent variable is valid and described with replicable precision 7 Dependent variables are measured repeatedly over time. 8 Data are collected on the reliability or interobserver agreement associated with each dependent variable, and IOA levels meet minimal standards (e.g., IOA = 80%; Kappa = 60%).
Independent variable	9 Independent variable is described with replicable precision 10 Independent variable is systematically manipulated and under the control of the experimenter. 11 Overt measurement of the fidelity of implementation for the independent variable is highly desirable.
Baseline	12 The majority of single-subject research studies will include a baseline phase that provides repeated measurement of a dependent variable and establishes a pattern of responding that can be used to predict the pattern of future performance, if introduction or manipulation of the independent variable did not occur. 13 Baseline conditions are described with replicable precision.
Experimental control/ internal validity	14 The design provides at least three demonstrations of experimental effect at three different points in time 15 The design controls for common threats to internal validity (e.g., permits elimination of rival hypotheses). 16 The results document a pattern that demonstrates experimental control.
External validity	17 Experimental effects are replicated across participants, settings, or materials to establish external validity.
Social validity	18 The dependent variable is socially important. 19 The magnitude of change in the dependent variable resulting from the intervention is socially important

	20 Implementation of the independent variable is practical and cost effective. 21 Social validity is enhanced by implementation of the independent variable over extended time periods, by typical intervention agents, in typical physical and social contexts.

Appendix E- WoE B Rating System

The rating system for WoE B was taken from Petticrew and Roberts' (2003) rating system for 'effectiveness' type investigations. Their rating system was adapted from Muir Gray's (1996) rating system. Systematic review was removed as there are none on this topic.

Table E1

WoE B Rating system

Rating score	Study type
3	Randomised control trials
2	Quasi-experimental studies, cohort studies
1	Qualitative research, survey, case-control studies, non-experimental evaluations.

Appendix F- WoE C Rating Systems

Gough's (2007) WoE C is designed to be specific to the research question and therefore was created for this investigation. Criteria and rationale can be found below. The average of these criteria was used to give an overall score.

Table F1

WoE C Rating system

Criteria	Weightings	Rationale
A Sample – grandparents	3 Grandparents included have significant caring responsibilities (11 or more waking hours a week on average) 2 Grandparents included have some caring responsibilities (between two and 11 waking hours a week on average) 1 Grandparents included have limited caring responsibilities (two waking hours a week or less on average) or caring responsibilities are not specified	Grandparents should have sufficient caring responsibilities to make an impact on a grandchild's life. 11 hours a week represents the equivalent of subsidised nursery places in the UK (Bradbury, Hoskins, & Fogarty (2021)
B Grandparent intervention	3 Intervention specifically tailored to and solely delivered to grandparents 2 Intervention either tailored specifically or solely delivered to grandparents 1 Intervention focused on grandparents but not specifically tailored or delivered to them	Intervention should be focused on grandparents in caregiving roles, preferably both in tailoring or in delivery
C Sample-Grandchildren	3 All grandchildren have a behaviour problem, measured by a standardised measure 2 All grandchildren have a behaviour problem identified by grandparents, teachers or parents 1 Grandchildren with no behaviour problems or with no details on how the behaviour problems were identified or not all grandchildren have behavioural problems	Grandchildren should have problem behaviour that is targeted and clearly measured
D Sample-grandchildren's development	3 Typically developing grandchildren with a mean age ± one standard deviation between zero and 12 years. 2 Typically developing grandchildren with a mean age ± one standard deviation between zero and 12 years from a specific group (e.g. ability or skill) or inclusion criteria on group is not specified (e.g. no mention if the	To ensure there is enough similarity of population to compare across samples

	group has been recruited from a particular group) 1 Grandchildren with a mean age ± one standard deviation outside zero and 12 years.	
E Measures of child behaviour	3 Grandparent, teacher and parent report 2 Reports from grandparent and teacher or parent 1 Reports from grandparent or teacher or parent	Triangulation of the measures is preferable
F Triple P Alignment	3 Triple P alone 2 Triple P with another aspect of an intervention 1 Minimal alignment with Triple P	It is preferable for studies to examine Triple P alone to ensure fidelity
G Follow up	3 Pre, post and at least six months check up 2 Pre, post data 1 Post data	Ideally, post and follow up data should be available to examine maintenance of any changes

Appendix G- Full breakdown of WoE scores

Table G1

WoE A- Gersten et al.'s (2005) quality of evidence indicators used with the RCTs

	Essential Criteria (see table B1)										Essential criteria met	Desirable criteria (see table B1)								Desirable criteria met	Overall WoE Score
	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10		D1	D2	D3	D4	D5	D6	D7	D8		
Hoang et al. (2022)	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	9	✓	X	✓	X	X	✓	X	✓	4	3
Kirby and Sanders (2014)	X	✓	X	✓	X	✓	✓	✓	✓	✓	7	✓	X	✓	X	X	✓	X	✓	4	1
Leung et al. (2014)	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	9	✓	✓	X	✓	✓	✓	X	✓	6	3
Smith et al. (2018)	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	9	X	X	✓	X	✓	✓	X	✓	4	3

Table G2

WoE A- Horner et al.'s (2005) quality of evidence indicators used with Kirby and Sanders' (2013) paper

Category	Criteria (see table B2)	Met?	Percent of criteria met for category	Score for this category
Description of Participants and Settings	1	Yes	33%	1
	2	No		
	3	No		
Dependent Variable	4	Yes	100%	3
	5	Yes		
	6	Yes		
	7	Yes		
	8	Yes		
Independent variable	9	Yes	67%	2
	10	Yes		
	11	No		
Baseline	12	No	50%	2
	13	Yes		
Experimental control/ internal validity	14	No	33%	1
	15	No		
	16	Yes		
External validity	17	No	0%	0
Social validity	18	Yes	75%	3
	19	No		
	20	Yes		
	21	Yes		

Note. overall score is 1.7

Table G3

WoE B with studies

Rating score	Study type	WoE B Score
Hoang et al. (2022)	Randomised control trial	3
Kirby and Sanders (2013)	Case study	1
Kirby and Sanders (2014)	Randomised control trial	3
Leung et al. (2014)	Randomised control trial	3
Smith et al. (2018)	Randomised control trial	3

Table G4

WoE C with study criteria

	A Sample – grandparents	B Grandparent intervention	C Sample-Grandchildren	D Sample-grandchildren's development	E Measures of child behaviour	F Triple P Alignment	G Follow up	Average
Hoang et al. (2022)	3	1	*2*	2	1	2	3	2.00
Kirby and Sanders (2013)	3	3	3	3	2	3	3	2.86
Kirby and Sanders (2014)	3	3	1	3	2	3	3	2.57
Leung et al. (2014)	3	3	1	2	1	3	3	2.29
Smith et al. (2018)	3	2	1	2	1	3	3	2.14