

## ***Case Study 1: An Evidence-Based Practice Review Report***

### ***How effective is the '1-2-3 Magic' parenting intervention at reducing child problem behaviours in British children in a child protection setting?***

#### **1. Summary**

There are Local Authorities in the United Kingdom that are currently funding training in the 1-2-3 Magic parenting intervention for caregivers with children with child in need, or child protection status. This intervention looks to reduce child problem behaviours by improving the caregiver-child relationship through principles based in cognitive-behavioural, operant-learning and social theory. Database searches found six randomised control trials (RCTs) addressing the effectiveness of this intervention, which were then evaluated for their methodological relevance and quality and their relevance to topic. Further exploration of the effect sizes were carried out alongside a meta-analysis.

Each paper reported significant positive effects of the intervention, although further analysis showed wide variation in effect size. Meta-analysis revealed a pooled positive effect of the intervention in reducing the perceived presence of problem behaviours but not on reducing the intensity of these behaviours. As the studies largely focussed on children outside a child protection setting, conclusions about the effectiveness of this intervention within that subpopulation cannot be made, however within samples representing the general population there were found to be positive, if varied, results.

## **2. Introduction**

### **2.1: What is 1-2-3 Magic?**

Brief parenting interventions are widely favoured due to their lower cost and ease of delivery compared to therapist-lead direct interventions (Cunningham, Bremner, & Boyle, 1995). One brief parenting intervention currently in practice worldwide is the '1-2-3 Magic' intervention stemming originally from Phelan's (1990) work, and later adapted by Phelan, (2010). This intervention is distinct to the '1-2-3 Magic and Emotion Coaching' approach (Hawton & Martin, 2011), as it does not draw on Emotion Coaching. In 2016 1-2-3 Magic had sold 1.5 million copies and had been translated into 22 languages, claiming to be the number 1 selling child discipline program in the US.

The 1-2-3 Magic programme looks to develop three main areas: stopping “obnoxious<sup>1</sup>” child behaviours; eliciting “positive<sup>1</sup>” child behaviours; and enhancing the relationship between caregiver and child. It does this through the application of seven key concepts which include caregivers (i) stating expectations clearly; (ii) consistently using the 1-2-3 counting systems as a warning for disruptive behaviours; (iii) consistently using time-out or equivalents; (iv) modelling self-regulation of emotion by minimising shouting and arguing; (v) praising positive behaviours; (vi) reassessing beliefs on child responses being logical; (vii) and removing anger from discipline in order to enhance the caregiver-child relationship.

---

<sup>1</sup> The terms “obnoxious” and “positive” are taken from Phelan (2010) and are widely used throughout delivery of the intervention.

Materials to support and inform caregivers in delivering the intervention can be purchased on an individual basis, and caregivers can self-deliver their own training using two published DVDs with a total runtime of under four hours (Booth & Phelan, 2004a, 2004b). However, training in the intervention is also offered in three 2-hour group sessions where caregivers can consolidate and refine newly learnt skills through discussion, role play, observation of “real world” application of techniques and individual feedback from the leading practitioner.

## **2.2: Psychological Basis for 1-2-3 Magic**

1-2-3 Magic draws heavily from three theoretical backgrounds. Perhaps the most explicitly expressed of these is Operant Learning Theory (Skinner, 1963), where the learner alters their behaviour based on external stimuli, drawing inference between a behaviour and related external consequence, this occurs until the new behaviour has been learnt. Examples within this intervention are the use of positive reinforcements when the child demonstrates positive behaviours in addition to the use of punishments, such as a time-out, when the child expresses negative behaviours.

This is further strengthened by the use of Cognitive-Behavioural Theory and Social Learning Theory within the intervention. Cognitive-Behavioural Theory states that negative behaviour patterns are a product of distorted thinking and that for behaviour to change these thought patterns must first be redefined. This is seen within the seven core principles where caregivers are trained and practise restructuring their beliefs that children respond to logical reasoning. Finally, Social Learning Theory (Bandura, 1978) – that we learn our behaviour through observation and imitation of those around us – is shown in this

parenting intervention as it trains parents to model emotional self-regulation by communicating in a non-argumentative way, and by removing anger from any disciplinary interactions.

### **2.3: Rationale for review**

Parenting style, attitude and behaviours are directly related to children's childhood behaviour outcomes (Smith & Farrington, 2004). With parenting interventions being seen to improve children's behaviour (Bor, Sanders, & Markie-Dadds, 2002; Bradley et al., 2003; Sanders, Markie-Dadds, Tully, & Bor, 2000), and to reduce levels of "clinical" disruptive behaviour in children (Bradley et al., 2003). Problematic and disruptive behaviours are found to significantly impact both children and caregivers (Flaherty, Stirling, & The Committee on Child Abuse and Neglect, 2010), and are exhibited by children from all parental income groups (Scott, 2008), which means the impact of parenting interventions can be extremely wide.

Although problem behaviours impact on the child in the moment they have also been found to have a long-term impact. Indeed, research has shown that children who demonstrate persistent problem behaviours have an increased risk for later unemployment, mental health disorders, anti-social behaviour and criminality (Bayer et al., 2011; Stevenson & Goodman, 2001). As addressing child problem behaviours through effective parenting interventions can theoretically have positive short and long-term effects it follows that many Local Authorities in the UK offer parenting interventions. The 1-2-3 Magic parenting intervention is currently offered by Bedfordshire, Northamptonshire, West Cheshire, Surrey and Kingston and Richmond,

amongst others, either as part of their local offer or by charities working within the area.

Local Authorities offer these funded courses to families most in need of support. The rationale behind this funding decision is largely based on policy surrounding early childhood disadvantage (Department for Education, 2011) and the research showing the long-term implications for this (Waldfogel & Washbrook, 2008). Often referrals for funded places are made by social services for children with a child in need, or child protection status, and therefore it is paramount that any intervention delivered to these specific groups have been shown to be effective amongst similar populations. In addition to this it is vital that Educational Psychologists have a full understanding of the parent training that the families they are working with have received, so they can ascertain whether influences on the child's behaviour from their home environment are likely to have changed.

#### **2.4: Review question**

How effective is the '1-2-3 Magic' parenting intervention at reducing child problem behaviours in British children in a child protection setting?

### 3. Critical Review of Evidence Base

#### 3.1: Systematic Literature Search

A comprehensive literature search was conducted in December 2018, using 5 electronic databases; ERIC (EBSCO), PsycInfo, PubMed, SCOPUS, and Web of Science. SCOPUS and Web of Science were chosen for their large selection of available publications; ERIC was chosen for its focus on Education; PsycInfo was chosen for its focus on psychology; and PubMed was chosen for its medical focus, in case the intervention has been used with a clinical population. Table 1 contains the search terms that were used to search for articles within these databases.

Table 1  
*Search Terms Used for Searching Databases*

Population	Intervention	Outcomes
“Nursery”	“1-2-3 Magic”	“Behaviour”
OR		OR
“Early Years”		“Behavior”
OR		
“Pre-School”		
OR		
“Child”		

Younger children were focussed on as part of this review, as evidence suggests that younger children show a greater improvement in conduct behaviour problems when their parents undergo a brief parenting intervention (Gardner, Hutchings, Bywater, & Whitaker, 2010).

The search identified 34 papers, whilst two further articles were found through the '1-2-3 Magic' website. An ancestral search identified a further three papers and research cited on the '1-2-3 Magic' website identified a further two, bringing the total number of articles to 39. 12 duplicates were then removed, and the remaining 27 articles underwent title screening followed by abstract screening using the inclusion and exclusion criteria outlined in Table 2. Figure 1 illustrates the literature search and Table 3 lists the included studies. Appendix 1 lists studies excluded at title and abstract screening.

Table 2  
*Inclusion and Exclusion Criteria for Screening Papers*

	Inclusion Criteria	Exclusion Criteria	Rationale
1 Language	Study is written in or translated into English.	Study is not written or translated into English.	To ensure research is correctly interpreted.
2 Participants and setting	Study participants are primary caregiver of child, who live with the child.	Study participants are not primary caregiver of child or live outside the child's home.	To ensure children are frequently exposed to the intervention.
	Study participants children have not been diagnosed with a behavioural disorder.	Study participants' children have been previously diagnosed with a behavioural disorder.	The review question focuses on a non-clinical group.
3 Type of Publication	Study is published in a peer-reviewed journal.	Study is unpublished or published in an unreviewed journal.	Peer reviewed journals contain a higher quality of studies with fewer errors.
	Study contains primary data.	Study is a systematic review or meta-analysis, containing secondary data.	Possible important information may be lost if the study has already been reviewed.

Table 2  
*Inclusion and Exclusion Criteria for Screening Papers*

	Inclusion Criteria	Exclusion Criteria	Rationale
4 Intervention	Study uses '1-2-3 Magic' intervention.	Study does not use '1-2-3 Magic' intervention.	This intervention is the focus of the review question.
5 Study Design	Study uses a Randomised Control Design, reporting pre- and post- quantitative data for an intervention and control group.	Study uses a qualitative or less stringent quantitative experimental design.	To ensure comparisons can be made between groups, and effect sizes can be generated in analysis.

Figure 1  
*Flow Diagram of Literature Search Process*

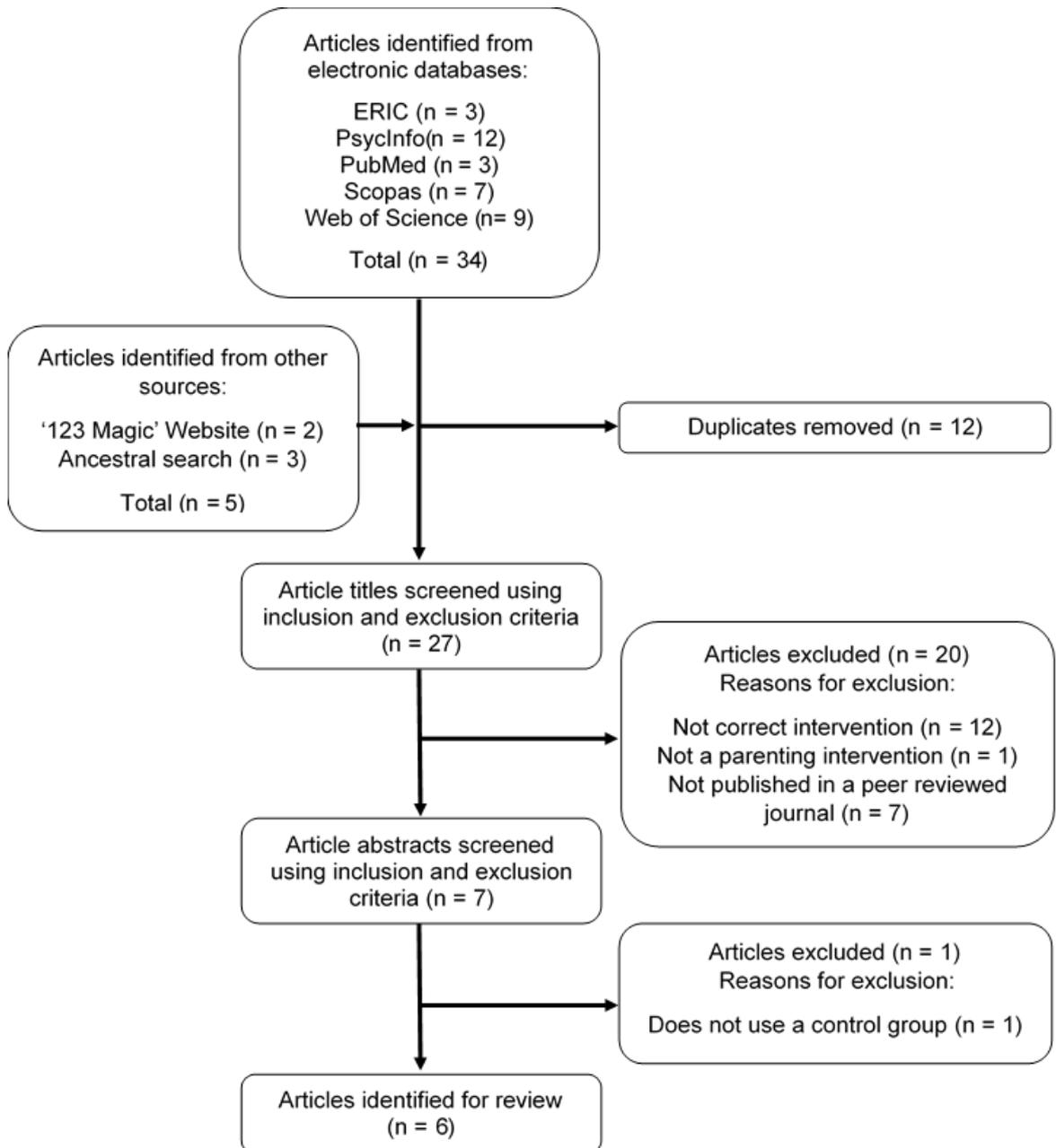


Table 3  
*Studies Included in this Review*

Study Code	Full Study Reference
A	Bailey, E. L., van der Zwan, R., Phelan, T. W., & Brooks, A. (2012). The 1-2-3 Magic Program: implementation outcomes of an Australian pilot evaluation with school-aged children. <i>Child &amp; Family Behavior Therapy, 34</i> (1), 53-69.
B	Bailey, E. L., Van Der Zwan, R., Phelan, T. W., & Brooks, A. (2015). Keeping it going: evidence of long-term improvements after implementation of the 1-2-3 Magic Parenting Program. <i>Child &amp; Family Behavior Therapy, 37</i> (4), 303-320.
C	Bradley, S. J., Jadaa, D. A., Brody, J., Landy, S., Tallett, S. E., Watson, W., & Stephens, D. (2003). Brief psychoeducational parenting program: An evaluation and 1-year follow-up. <i>Journal of the American Academy of Child &amp; Adolescent Psychiatry, 42</i> (10), 1171-1178.
D	Porzig-Drummond, R., Stevenson, R. J., & Stevenson, C. (2014). The 1-2-3 Magic parenting program and its effect on child problem behaviors and dysfunctional parenting: A randomized controlled trial. <i>Behaviour research and therapy, 58</i> , 52-64.
E	Porzig-Drummond, R., Stevenson, R. J., & Stevenson, C. (2015). Preliminary evaluation of a self-directed video-based 1-2-3 Magic parenting program: A randomized controlled trial. <i>Behaviour research and therapy, 66</i> , 32-42.
F	Flaherty, R., & Cooper, R. (2010). Piloting a parenting skills program in an Australian rural child protection setting. <i>Children Australia, 35</i> (3), 18-24.

### 3.2: Mapping the Field

By using a systematic literature search six studies have been identified that describe the effects on behavioural outcomes of children whose caregivers have undergone training in the '1-2-3 Magic' parenting intervention. Table 4 describes the key features for each of these studies.

Table 4  
*Key Features of Each Study*

Study and Code	Study Design	Study Location	Sample Characteristics	Intervention Details	Measures	Outcomes and Key Findings
[A] Bailey et al. (2012)	Randomised control trial using wait-list control.	Australia	9 families Children aged 6-12 Male = 5 Female = 4 Mid-High SES	2 3-hour sessions lead by a trainer, focussing on 'stop' and 'start' behaviours.	Child behaviour measured using Intensity and Problem scales on Eyberg Child behaviour Inventory (ECBI)	Post-intervention: Significant improvement was seen in intensity and problem scales of intervention group but not seen in control group. However, no comparisons were made between the two groups.
[B] Bailey et al. (2015)	Randomised control trial using wait-list control.	Australia	12 families Children aged 6-12 Male = 9 Female = 3 Mid-High SES	2 3-hour active skill sessions lead by a trainer on 'stop' and 'start' behaviours.	Child behaviour measured using Intensity and Problem scales on Eyberg Child behaviour Inventory (ECBI)	Post-intervention: Significant reduction in intensity and problem behaviours seen in intervention but not control group. No comparisons made between groups. 1 year: Significant reduction in intensity and problem scales compared to baseline.

Table 4

*Key Features of Each Study*

Study and Code	Study Design	Study Location	Sample Characteristics	Intervention Details	Measures	Outcomes and Key Findings
[C] Bradley et al. (2003)	Randomised control trial using wait-list control.	Toronto, Canada Urban area	198 families Children ages 3-4 Male = 121 Female = 77 Mid-High SES	3 2-hour group meetings followed by a booster session 4 weeks later.	Child behaviour measured using Aggressive, Anxious, Hyperactive scales of Preschool Behaviour Questionnaire (PBQ)	Post-intervention: Significant improvements were made in all 4 scales of child behaviour from before intervention. Significance was also seen between control and intervention group. 1 year: no significant change from post-intervention.
[D] Porzig-Drummond et al. (2014)	Randomised control trial using wait-list control.	Sydney, Australia	92 families Children aged 2-12 Male = 54 Female = 38 Mid-High SES	3 x 2hour sessions including DVD viewing, follow up Q&A session. Lead by psychologist	Child behaviour measured using Intensity and Problem scales on Eyberg Child behaviour Inventory (ECBI)	Post-intervention: significant reduction on intensity and problem scales and significant differences between control and intervention group 3 months: no significant change from post-intervention. 2 year: no significant change from post-intervention.

Table 4  
*Key Features of Each Study*

Study and Code	Study Design	Study Location	Sample Characteristics	Intervention Details	Measures	Outcomes and Key Findings
[E] Porzig-Drummond et al. (2015)	Randomised control trial using wait-list control.	New South Wales, Australia Rural and urban areas	84 families Children ages 2-10 Male = 42 Female = 42 Mid-High SES	DVDs are viewed at home with guidance and Q&A accessible via email. Questionnaires are also completed online.	Child behaviour measured using Intensity and Problem scales on Eyberg Child behaviour Inventory (ECBI)	Post-intervention: significant reduction on intensity and problem scales and significant differences between control and intervention group 6 months: no significant change from post-intervention.
[F] Flaherty and Cooper (2010)	Randomised control trial using wait-list control.	New South Wales, Australia. Rural area.	35 families (99 children). No further sample characteristic data provided.	3 sessions of unstated length were given to each caregiver.	Child behaviour measured using Intensity and Problem scales on Eyberg Child behaviour Inventory (ECBI)	Post-intervention: significant differences were found between pre- and post-intervention ratings of child behaviour in the intervention group.

### **3.3: Weight of Evidence (WoE)**

To evaluate each of the six studies Gough's (2007) weight of evidence (WoE) framework was used. This examines three main areas within each study: methodological quality (WoE A), methodological relevance (WoE B), and the relevance of the study to the review question (WoE C).

WoE A was assessed using an adapted version of Law et al.'s (1998) coding protocol for quantitative studies. Sections not pertaining to the methodological quality of the paper (such as study purpose, and quality of background literature) were not used as this would not inform WoE A. Additional information for WoE A can be seen in Appendix 2.

To examine the appropriateness of each study for answering this review question (WoE B) a typology of evidence was applied (Petticrew & Roberts, 2003). As all studies are randomised control trials a further distinction was applied using the PEDro Scale (Maher, Sherrington, Herbert, & Moseley, 2003). WoE C evaluates the relevance of the studies for answering the review question. Appendices 3 and 4 contain additional information about WoE B and C.

Collectively WoE A, B and C allow an overall weight of evidence to be rated (WoE D). Judgments pertaining to weight of evidence can be seen in Table 5, with additional supporting information and example coding protocols in Appendices 5 and 6 respectively.

Table 5  
*Summary of Weight of Evidence Judgements*

Studies	WoE A: Methodological Quality	WoE B: Methodological Relevance	WoE C: Relevance to Topic	WoE D: Overall Score
<b>A</b> Bailey et al. (2012)	1.8 (Low)	2 (Medium)	2 (Medium)	5.8 (Medium)
<b>B</b> Bailey et al. (2015)	1.4 (Low)	2 (Medium)	2 (Medium)	5.4 (Medium)
<b>C</b> Bradley et al. (2003)	1.8 (Low)	2 (Medium)	1.75 (Low)	5.55 (Medium)
<b>D</b> Porzig- Drummond et al. (2014)	2.2 (Medium)	3 (High)	2 (Medium)	7.2 (High)
<b>E</b> Porzig- Drummond et al. (2015)	2.2 (Medium)	2 (Medium)	1 (Low)	5.2 (Medium)
<b>F</b> Flaherty and Cooper (2010)	1.2 (Low)	1 (Low)	1.25 (Low)	3.45 (Low)

### 3.4: Participants

In total the 430 caregivers received instruction in 1-2-3 Magic as part of an intervention group. Five studies asked caregivers to report on the behaviour of just one child, with one study (Flaherty & Cooper, 2010) asking for an estimation of the average effect of the intervention across all their children. There was variation seen in the number of participants in each study ranging from nine (Bailey, van der Zwan, Phelan, & Brooks, 2012) to 198 (Bradley et al., 2003). Ratings for determining statistical power are reported as part of WoE A, showing that one paper calculated the sample size required to achieve good statistical power (Porzig-Drummond, Stevenson, & Stevenson, 2014), although they were unable to retain enough participants in their post-

intervention measures to ensure this sample size was met. There was variation in ratings for WoE A based on sample selection and description, with Porzig-Drummond et al., (2014) and Porzig-Drummond et al., (2015) receiving a high rating due to the justification of sample size, Flaherty & Cooper (2010), receiving a low rating due to not comparing difference in control and intervention groups prior to intervention. The remaining papers (Bailey et al., 2012; Bailey, van der Zwan, Phelan & Brooks, 2015), received a medium rating as they compared the control and intervention groups prior to intervention but did not give statistical justification of their chosen sample size.

Collectively, caregivers were asked to report on children ranging in age from 2-16 years old. Bradley et al., (2003) had the smallest age range of 3-4 year olds whilst the other studies reported on children aged, 6-12 (Bailey et al., 2012; Bailey, van der Zwan, Phelan & Brooks, 2015), 2-12 (Porzig-Drummond et al., 2014), 2-10 (Porzig-Drummond, Stevenson, & Stevenson, 2015) and 2-16 years (Flaherty & Cooper, 2010). Bailey et al., (2012) and Bailey et al., (2015) specifically asked caregivers to report on children without a SEND diagnosis of any kind, whilst Porzig-Drummond et al., (2014) and Porzig-Drummond et al., (2015) excluded caregivers of parents with diagnosed behavioural disorders such as ADHD. The remaining two studies did not report on SEND of their participants.

All studies were conducted in westernised countries, one in Canada (Bradley et al, 2003) and all others in Australia (Bailey et al., 2012; Bailey et al., 2015; Porzig-Drummond et al., 2014; Porzig-Drummond et al., 2015; Flaherty & Cooper, 2010). All studies reported the socioeconomic status (SES) of their participants as mid-high, apart from Flaherty and Cooper (2010) who did not

report SES status. Their sample also varied from the other sample groups as all participating families had been exposed or involved in domestic abuse which in the UK would equate to a Child in Need or Child Protection setting, this study therefore was the only study receiving a medium weighting for population as part of WoE C.

### **3.5: Research Design**

All studies selected were Randomised Control Trials as this study design has the highest methodological rigour (Petticrew & Roberts, 2003). This meant that every study collected data from two groups of participants – intervention and waitlist control group - at pre- and post-intervention time points. Four studies also collected data at follow up points ranging from three months to two years (Bailey et al., 2015; Bradley et al., 2003; Porzig-Drummond et al., 2014; Porzig-Drummond et al., 2015). Every study, apart from Flaherty and Cooper (2010), checked for equivalence between the control and intervention groups.

Randomisation of participants was described by Porzig-Drummond et al., (2014) and Porzig-Drummond et al., (2015) as using Excel-generated randomisation schedules, however no other study described their randomisation process. This is coded as part of WoE B, where five studies (Bailey et al., 2012; Bailey et al., 2015; Bradley et al., 2003; Porzig-Drummond et al., 2014; Porzig-Drummond et al., 2015) received a higher rating on the PEDro scale as they were able to match their control and intervention groups at baseline. The highest scoring paper for WoE B (Porzig-Drummond et al., 2014) was additionally able to blind all subject by providing alternative intervention and provided explanation for how the subjects were randomly allocated.

### **3.6: Intervention**

Whilst all studies report using the 1-2-3 Magic parenting intervention, there is much variation in delivery. The importance of this is reflected in the coding of intervention delivery within both WoE A and C. Two studies delivered the recommended six hours of training scoring a medium rating in WoE C (Bailey et al., 2012; and Bailey et al., 2015), two studies delivered additional training on top of the recommended six hours scoring a high rating in WoE C (Bradley et al., 2003, and Porzig-Drummond et al., 2014). However low WoE C rating were given to Porzig-Drummond et al., (2015) as they exposed participants to training via a posted DVD and Flaherty and Copper (2010) who did not state how participants received the intervention. There was also variation in who was delivering the intervention ranging from a trained psychologist (Porzig-Drummond et al., 2014), or trained non-psychologist training (Bailey et al., 2012; Bailey et al., 2015, and Bradley et al., 2003), to self-administration (Porzig-Drummond et al., 2015). Again, Flaherty and Cooper (2010), did not describe the delivery of the intervention in detail.

### **3.7: Measures**

All studies used a caregiver questionnaire response to measure the child's behaviour, with all studies using the Eyberg Child behaviour Inventory (ECBI) (Eyberg & Pincus, 1999) apart from Bradley et al., (2003) who used the Preschool Behaviour Questionnaire (Behar & Stringfield, 1974). This meant that type of problem behaviour was measured in Bradley et al., (2003), whereas all other studies measured intensity of child problem behaviours, alongside the presence of generic 'problem behaviours' displayed. Every study

also measured parent outcomes, but these have been excluded as they are not part of this review question.

### **3.8: Outcomes and Effect Sizes**

All six studies reported positive effects of the 1-2-3 Magic intervention on reducing problem behaviours in children. However, data-reporting was varied despite similar measurements being used to obtain pre- and post- intervention data for the intervention and control groups. Porzig-Drummond et al., (2014), Porzig-Drummond et al., (2015), and Bradley et al., (2003), each reported F-test results from ANOVAs which allowed for Pearson's  $r$  to be calculated from partial-eta squared ( $\eta_p^2$ ). Flaherty and Cooper (2010) reported means and standard error which meant Pearson's  $r$  could also be calculated. Bailey et al., (2012) and Bailey et al., (2015) only reported t-test scores comparing pre- and post-intervention data for the intervention group, and separately for the control group. As they did not use an ANOVA to make comparisons between the groups it is not possible to report an effect size.

Both Bailey et al., (2012) and Bailey et al., (2015) report significant improvements in reducing intensity and presence of problem behaviours, however due to their choice of statistical analysis (independent t-tests for the intervention and waitlist control rather than use of an ANOVA to also compare differences between groups) these results should be interpreted with caution.

Weak to moderate correlations were seen between implementation of 1-2-3 Magic and perceived presence of child problem behaviours (Bradley et al., 2003; Porzig-Drummond et al., 2015; Flaherty & Copper, 2010) and intensity of child problem behaviours (Porzig-Drummond et al., 2015; Flaherty &

Copper, 2010). Further correlation was shown by Porzig-Drummond et al., (2014) between implementation of the intervention and both intensity and presences of problem behaviours. This correlation was shown to be strong immediately following intervention, with intensity of problem behaviour continuing to show a weak positive correlation with the intervention at three months follow-up and the presence of problem behaviours showing a continued weak positive correlation at 2 years post-intervention. Calculated effect sizes can be seen in Table 6.

Table 6  
Summary of Effect Sizes

Studies and Code	Outcome Measure	Effect size (Pearson's r)						Participant Numbers	WoE D:
		T1 – post-intervention		T2 – follow-up		T3 – follow=up			
[A] Bailey et al. (2012)	Child behaviour: Intensity Problem	*		n/a		n/a		9	5.8 (Medium)
[B] Bailey et al. (2015)	Child behaviour: Intensity Problem	*		1 year **		n/a		12	5.4 (Medium)
[C] Bradley et al. (2003)	Child behaviour: Aggressive Anxious Hyperactive Total	r = 0.15 r = 0.27 r = 0.22 r = 0.15	weak weak weak weak	1 year **		n/a		198	5.55 (Medium)
[D] Porzig-Drummond et al. (2014)	Child behaviour: Intensity Problem	r = 0.56 r = 0.53	strong strong	3 months r = 0.15 r = 0.10	weak weak	2 years r = 0.00 r = 0.20	no effect weak	92	7.2 (High)
[E] Porzig-Drummond et al. (2015)	Child behaviour: Intensity Problem	r = 0.49 r = 0.42	moderate moderate	6 months **		n/a		84	5.2 (Medium)

[F] Flaherty and Cooper (2010)	Child behaviour: Intensity Problem	*** r = 0.10 r = 0.32	weak moderate	n/a	n/a	35	3.2 (Low)
--------------------------------	--	-----------------------------	------------------	-----	-----	----	--------------

---

.1 = weak, .3 = moderate, .5 = strong

\* effect size cannot be calculated as inappropriate statistical analysis was carried out and presented.

\*\* effect size cannot be calculated as required data not presented in study.

\*\*\* note that sample size was not stated so “worst-case” dropout rate has been assumed, with all 3 participants leaving from the same group. This results in a 19 in intervention group and 16 in control group.

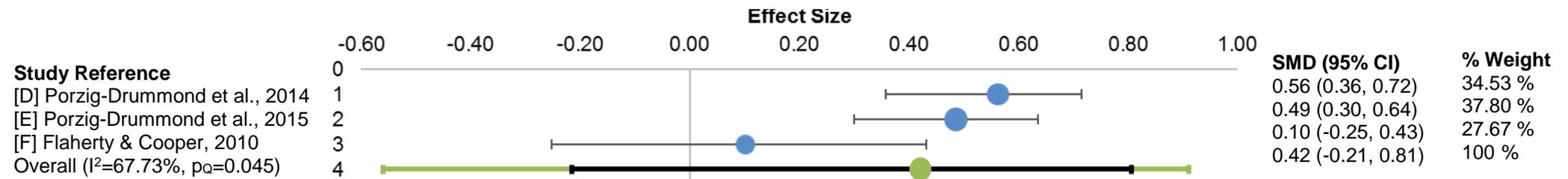
### 3.9: Meta-analysis

Data measuring the intensity of child problem behaviour was available for three studies (Porzig-Drummond et al., 2014; Porzig-Drummond et al., 2015; Flaherty & Cooper, 2010) immediately following intervention, reporting results of 180 parent responses in total. Despite two studies showing a positive effect of the intervention, the pooled analysis showed the intervention was not associated with a reduction in intensity of child problem behaviours ( $r = 0.42$ , 95% CI = -0.21, 0.81,  $p = 0.045$ ). There were significant values and substantial ranges of heterogeneity ( $I^2=67.73\%$ ,  $p_Q=0.045$ ), this represents a non-random variation with possible moderating effects.

Additional data regarding the perceived presence of problem behaviours displayed was also collected by the same three studies, in addition to this total frequency of problem behaviour. Data measuring the number of different problem behaviours were also collected by Bradley et al., (2003). Although this information was collected using a different measurement tool it is reporting on presence of problem behaviours. In total data reflecting the presence of problem behaviours in children was reported for 378 individuals. Here pooled analysis showed a positive association between implementing the parenting intervention and reducing the presence of child problem behaviours ( $r = 0.35$ , 95% CI = 0.06, 0.59,  $p = 0.012$ ). Significant heterogeneity was also detected within this measure ( $I^2=72.73\%$ ,  $p_Q=0.012$ ), showing further non-random variation with possible moderating effects. This information is represented in the Forest plots in Figure 2 and Figure 3.

Figure 2

Forest plot to show effect size, confidence intervals and percentage weighting for reduction in intensity of problem behaviours.



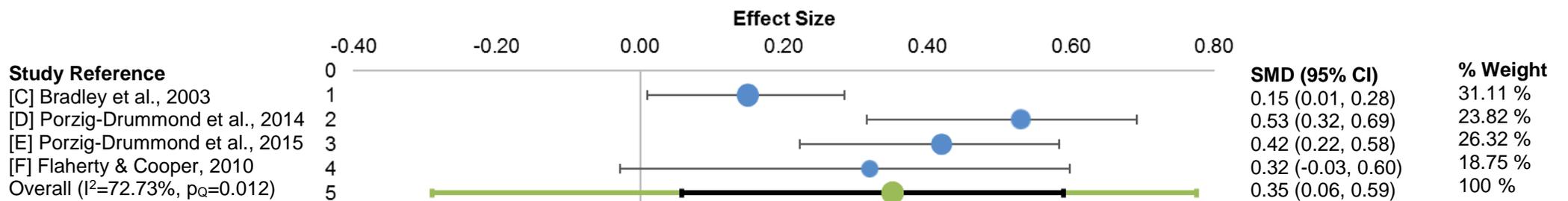
SMD = Standardised Mean Difference

CI = Confidence Intervals

Effect Size = Pearson's r

Figure 3

Forest plot to show effect size, confidence intervals and percentage weighting for reduction in problem behaviours.



SMD = Standardised Mean Difference

CI = Confidence Intervals

Effect Size = Pearson's r

## 4. Conclusion and Recommendations

This literature review looked to evaluate the effectiveness of the 1-2-3 Magic parenting intervention on improving child problem behaviours for children in a child protection setting. Research conducted within the child protection setting was limited to one paper (Flaherty & Cooper, 2010), which was included alongside five others which met the inclusion criteria. Use of Gough's (2007) weight of evidence (WoE) framework, alongside calculation of effect size has provided a mixed evidence base for the overall effectiveness of this intervention and revealed a gap in literature regarding its use in a child protection setting. WoE D ratings ranged from low (Flaherty & Cooper, 2010) to high (Porzig-Drummond et al., 2014) with all others receiving a medium rating. Effect sizes also showed a range from weak to strong for both intensity and perceived presence of problem behaviours, however pooled analysis only revealed an overall effect of 1-2-3 Magic on reducing the perceived presences of problem behaviours, not intensity, this was due to the weighting of studies based on their sample sizes.

The initial decision to only include RCTs was made due to the methodological rigour that is afforded by them, and whilst elements of each study show methodological rigour there are still flaws within each study. Only one study (Porzig-Drummond et al., 2014) assessed the impact of intervention on children within the full age range that the 1-2-3 Magic intervention is aimed at, and whilst only one study works with children outside the 2-12 suggested age range (Flaherty & Cooper, 2010) it is still pertinent to consider that any positive effects of the intervention noted in these children in the remaining four studies may not be applicable to children in the wider, 2-12, age range.

RCTs are the most favourable study design for looking at effectiveness of interventions (Petticrew & Roberts, 2003), but they are not well placed for looking at small and varied population groups such as children from a child protection setting. This may be one of the reasons that there is limited evidence for the effectiveness of this intervention within this population group. As researchers strive for methodological rigour they use RCTs which invariably need large sample sizes to ensure they are sufficiently powered, therefore working with a small and varied population is not easily achieved as part of an RCT. Whilst an RCT is desirable for looking at effectiveness of interventions there are alternative ways, for example a small-n experimental design. Therefore, to address the current gap in the literature regarding the effectiveness of the 1-2-3 Magic parenting intervention in a child protection setting, further research should be conducted with this population using alternative study designs.

As RCTs were used in each study, all studies benefitted from having wait-list control groups meaning that comparisons could be made between participants as well as within participants. However, none of the studies utilised an active control group – receiving alternative intervention or null intervention – participants may have been influenced in their post-intervention judgements of child behaviour due to their awareness of receiving the intervention. Alternatively, instead of asking caregivers to self-report on their children's behaviour, experimenter observations could have been made. The experimenter would be blinded to grouping during post-intervention observations therefore removing the social desirability effect of the parents

who received the intervention wanting to show an improvement in their child's behaviour.

Substantial heterogeneity was found in studies for both measurements of intensity and perceived presence of problem behaviours. Whilst this is not necessarily a bad thing it does suggest that there are moderators having an effect within the studies, and there is scope to explore these further. Additionally, the high heterogeneity, demonstrating inconsistency in the findings, provides further support for standardising studies more carefully. For example, ensuring fidelity of intervention during delivery, as two studies either failed to provide or were unable to evidence the full six-hour intervention being delivered (Porzig-Drummond et al., 2015; Flaherty & Cooper, 2010).

Although these studies provide insufficient evidence to suggest that child problem behaviours can be reduced in British children from a child protection setting using the 1-2-3 Magic parenting intervention, there is some evidence towards the effectiveness of the intervention in other populations. As some British Local Authorities are currently recommending the 1-2-3 Magic intervention to families referred to them through child protection, or child in need pathways, further research will need to be conducted in this population. Additionally, Educational Psychologists could consider acting as a critical friend within Local Authorities that run this parenting intervention, or at least retain a critical viewpoint on "improved" parenting techniques used within families who have received the 1-2-3 Magic parenting intervention.

## 5. References

- Bailey, E. L., van der Zwan, R., Phelan, T. W., & Brooks, A. (2012). *The 1-2-3 Magic Program: Implementation Outcomes of an Australian Pilot Evaluation With School-Aged Children*. *Child & Family Behavior Therapy, 34*(1), 53–69.
- Bailey, E. L., Van Der Zwan, R., Phelan, T. W., & Brooks, A. (2015). Keeping It Going: Evidence of Long-Term Improvements After Implementation of the 1-2-3 Magic Parenting Program.
- Bandura, A. (1978). Social Learning Theory of Aggression. *Journal of Communication, 28*(3), 12–29. Retrieved from <https://eric.ed.gov/?id=EJ195900>
- Bayer, J. K., Ukoumunne, O. C., Lucas, N., Wake, M., Scalzo, K., & Nicholson, J. M. (2011). Risk factors for childhood mental health symptoms: national longitudinal study of Australian children. *Pediatrics, 128*(4), e865-79.
- Behar, L., & Stringfield, S. (1974). A behavior rating scale for the preschool child. *Developmental Psychology, 10*(5), 601.
- Booth, E. C., & Phelan, T. W. (2004a). *1-2-3 Magic: Managing difficult behavior in children 2-12 [DVD]*. Glen Ellyn, IL: ParentMagic.
- Booth, E. C., & Phelan, T. W. (2004b). *More 1-2-3 Magic: Encouraging good behavior, independence and self-esteem [DVD]*. Glen Ellyn, IL: ParentMagic.
- Bor, W., Sanders, M. R., & Markie-Dadds, C. (2002). The Effects of the Triple P-Positive Parenting Program on Preschool Children with Co-Occurring Disruptive Behavior and Attentional/Hyperactive Difficulties. *Journal of Abnormal Child Psychology, 30*(6), 571–587.
- Bradley, S. J., Jadaa, D., Brody, J., Landy, S., Tallett, S. E., Watson, W., ... Stephens, D. (2003). Brief Psychoeducational Parenting Program: An Evaluation and 1-Year Follow-up. *Journal of American Academy Child and Adolescent Psychiatry, 42*(10), 1171–1178.
- Cunningham, C. E., Bremner, R., & Boyle, M. (1995). Large Group Community-Based Parenting Programs for Families of Preschoolers at Risk for Disruptive Behaviour Disorders: Utilization, Cost Effectiveness, and Outcome. *Journal of Child Psychology and Psychiatry, 36*(7), 1141–1159.
- Depart for Education. (2011). *A New Approach to Child Poverty: Tackling the Causes of Disadvantage and Transforming Families' Lives*. Retrieved from [www.official-documents.gov.uk](http://www.official-documents.gov.uk)
- Eyberg, S. M., & Pincus, D. (1999). *ECBI & SESBI-R: Eyberg child behavior inventory and Sutter-Eyberg student behavior inventory-revised: Professional manual*. Odessa, FL: Psychological Assessment Resources.

- Flaherty, E., Stirling, J., & The Committee on Child Abuse and Neglect. (2010). Clinical Report - The Pediatrician's Role in Child Maltreatment Prevention. *American Academy of Pediatrics*, 126(4), 833–841.
- Flaherty, R., & Cooper, R. (2010). Piloting a parenting skills program in an Australian rural child protection setting. *Children Australia*, 35(03), 18–24.
- Gardner, F., Hutchings, J., Bywater, T., & Whitaker, C. (2010). Who Benefits and How Does It Work? Moderators and Mediators of Outcome in an Effectiveness Trial of a Parenting Intervention. *Journal of Clinical Child & Adolescent Psychology*, 39(4), 568–580.
- Gough, D. (2007). Weight of Evidence: a framework for the appraisal of the quality and relevance of evidence.
- Hawton, M., & Martin, K. (2011). *1–2–3 Magic & Emotion Coaching Parenting Program 2011 (Power Point and manual)*.
- Law, M., Stewart, D., Pollock, N., Letts, L., Bosch, J., & Westmorland, M. (1998). *Guidelines for Critical Review Form - Quantitative Studies*. Retrieved from <http://www-fhs.mcmaster.ca/rehab/ebp/http://www-fhs.mcmaster.ca/rehab/ebp/pdf/qualguidelines.pdf><http://www-fhs.mcmaster.ca/rehab/ebp/pdf/qualreview.pdf>
- Maher, C., Sherrington, C., Herbert, R., & Moseley, A. (2003). Reliability of the PEDro Scale for Rating Quality of Randomized Controlled Trials. *Physical Therapy*.
- Petticrew, M., & Roberts, H. (2003). Evidence, hierarchies, and typologies: horses for courses. *Journal of Epidemiology and Community Health*, 57, 527–529.
- Phelan, T. W. (1990). *1-2-3 Magic: Training Your Preschoolers and Preteens to Do what You Want!* U.S.: Child Management Inc.
- Phelan, T. W. (2010). *1-2-3 magic*. ParentMagic. Retrieved from [https://books.google.co.uk/books?hl=en&lr=&id=ONHFAGAAQBAJ&oi=fnd&pg=PP1&dq=1-2-3+Magic:+Effective+Discipline+for+Children+2-12&ots=hIhSU-Fn\\_W&sig=EbYH76tKRSNdZhvR9cVYZwfpAPk#v=onepage&q=1-2-3+Magic%3A+Effective+Discipline+for+Children+2-12&f=false](https://books.google.co.uk/books?hl=en&lr=&id=ONHFAGAAQBAJ&oi=fnd&pg=PP1&dq=1-2-3+Magic:+Effective+Discipline+for+Children+2-12&ots=hIhSU-Fn_W&sig=EbYH76tKRSNdZhvR9cVYZwfpAPk#v=onepage&q=1-2-3+Magic%3A+Effective+Discipline+for+Children+2-12&f=false)
- Porzig-Drummond, R., Stevenson, R. J., & Stevenson, C. (2014). The 1-2-3 Magic parenting program and its effect on child problem behaviors and dysfunctional parenting: A randomized controlled trial. *Behaviour Research and Therapy*, 58, 52-64
- Porzig-Drummond, R., Stevenson, R. J., & Stevenson, C. (2015). Preliminary evaluation of a self-directed video-based 1-2-3 Magic parenting program: A randomized controlled trial. *Behaviour Research and Therapy*, 66, 32–42.
- Sanders, M. R., Markie-Dadds, C., Tully, L. A., & Bor, W. (2000). The Triple P-Positive Parenting Program: A Comparison of Enhanced, Standard,

and Self-Directed Behavioral Family Intervention for Parents of Children With Early Onset Conduct Problems. *Journal of Consulting and Clinical Psychology*, 68, 624-640. . Retrieved from [https://s3.amazonaws.com/academia.edu.documents/41899957/Sanders\\_MR\\_Markie-Dadds\\_C\\_Tully\\_LA\\_Bor\\_W20160202-11441-2ltsv5.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1547394732&Signature=a92AINhiv4b%2Be04inQeRclHy0w8%3D&response-content-disposition=inline%3Bfilename%3DThe\\_Triple\\_P-Positive\\_Parenting\\_Program.pdf](https://s3.amazonaws.com/academia.edu.documents/41899957/Sanders_MR_Markie-Dadds_C_Tully_LA_Bor_W20160202-11441-2ltsv5.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1547394732&Signature=a92AINhiv4b%2Be04inQeRclHy0w8%3D&response-content-disposition=inline%3Bfilename%3DThe_Triple_P-Positive_Parenting_Program.pdf)

- Scott, S. (2008). Parenting programmes for attachment and conduct problems. *Psychiatry*, 7(9), 367–370.
- Skinner, B. F. (1963). Operant behavior. *American Psychologist*, 18(8), 503–515. <https://doi.org/10.1037/h0045185>
- Smith, C. A., & Farrington, D. P. (2004). Continuities in antisocial behavior and parenting across three generations. *Journal of Child Psychology and Psychiatry*, 45(2), 230–247.
- Stevenson, J., & Goodman, R. (2001). Association between behaviour at age 3 years and adult criminality. *British Journal of Psychiatry*, 179(03), 197–202.
- Waldfoegel, J., & Washbrook, E. (2008). ). “Early years policy”. In Social Mobility and Education: Academic papers presented a high level summit sponsored by the Carnegie Corporation of New York and the Sutton Trust, 1-3 June 2008.

# 1. Appendices

## Appendix 1 – Excluded Studies

Table 1  
*List of Excluded Studies*

Stage	Full Study Reference	Reason for Exclusion
Title	Bunge. (2016). Behavioral intervention technologies and psychotherapy with youth: A review of the literature. <i>Current Psychiatry Reviews</i> , 12(1), 14-28.	4
Title	Gulliford. (2015). Teaching coping skills in the context of positive parenting within a preschool setting. <i>Australian Psychologist</i> , 50(3), 219-231.	4
Title	Perez, E., Turner, M., Fisher, A., Lockwood, J., & Daley, D. (2014). Linguistic analysis of the Preschool Five Minute Speech Sample: what the parents of preschool children with early signs of ADHD say and how they say it?. <i>PloS one</i> , 9(9), e106231.	4
Title	Benzies, K., Mychasiuk, R., Kurilova, J., Tough, S., Edwards, N., & Donnelly, C. (2014). Two-generation preschool programme: Immediate and 7-year-old outcomes for low-income children and their parents. <i>Child &amp; Family Social Work</i> , 19(2), 203-214.	4
Title	Kendall, S., Bloomfield, L., Appleton, J., & Kitaoka, K. (2013). Efficacy of a group-based parenting program on stress and self-efficacy among Japanese mothers: A quasi-experimental study. <i>Nursing &amp; health sciences</i> , 15(4), 454-460.	4
Title	Schneider. (2012). <i>The science of consequences: How they affect genes, change the brain, and impact our world.</i>	4
Title	Landy. (2009). <i>Pathways to competence: Encouraging healthy social and emotional development in young children.</i>	4
Title	Porter. (2008). <i>Young children's behavior: Practical approaches for caregivers and teachers.</i>	4
Title	Kristal. (2005). <i>The temperament perspective: Working with children's behavioral styles.</i>	4
Title	Allen, S. M., Thompson, R. H., & Drapeaux, J. (1997). Successful methods for increasing and improving parent and child interactions. <i>NHSA Dialog: A Research-to-Practice Journal for the Early Intervention Field</i> , 1(3), 58-80.	4
Title	Diana, D. P. (2010). <i>Marketing for the Mental Health Professional: An Innovative Guide for Practitioners.</i> John Wiley & Sons.	4
Title	Schwartz. (2004). <i>The new language of toys: Teaching communication skills to children with special needs.</i>	4

Table 1  
*List of Excluded Studies*

Stage	Full Study Reference	Reason for Exclusion
Title	Barton, C. L. (2013). <i>Effectiveness of 1-2-3 Magic for Teachers to improve classroom management and discipline</i> (Doctoral dissertation, Capella University).	2a
Title	Smith, J. (2017). 1-2-3 Magic for Kids: Helping Your Kids Understand the New Rules.	3a
Title	Smith, J. (2016). 1-2-3 Magic Teen: Communicate, Connect, and Guide Your Teen to Adulthood.	3a
Title	Phelan, T. W. (2010). <i>1-2-3 Magic: Effective Discipline for Children 2–12</i> . ParentMagic, Inc..	3a
Title	Summary Toronto Hospital for Sick Children Study (2014). Retrieved from: <a href="https://www.123magic.com/parenting-tips/summary-toronto-hospital-for-sick-children-study.html">https://www.123magic.com/parenting-tips/summary-toronto-hospital-for-sick-children-study.html</a>	3a
Title	Mooseheart Child City and School Study (2003). Retrieved from: <a href="https://www.123magic.com/parenting-tips/mooseheart-child-city-and-school-study.html">https://www.123magic.com/parenting-tips/mooseheart-child-city-and-school-study.html</a>	3a
Title	Bloomfield, L., & Kendall, S. (2010). Audit as evidence: the effectiveness of '123 magic' programmes. <i>Community Practitioner</i> .	3a
Title	Bailey, E. L., van der Zwan, R., Phelan, T. W., & Brooks, A. (2009). The 1–2-3 Magic Program: An evaluation of a brief psychoeducational parenting program. In N. Voudouris & V. Mrowski (Eds.), <i>Proceedings of the 44th Annual Australian Psychological Society Conference</i> (pp. 7–12). Melbourne, Australia: The Australian Psychological Society Ltd.	3a
Abstract	Porzig-Drummond, R., Stevenson, R. J., & Stevenson, C. (2016). A Preliminary Evaluation of the 1-2-3-Magic Parenting Program in an Australian Community Services Setting. <i>Australian Social Work</i> , 69(4), 388-402.	5

## Appendix 2 – WoE A (Methodological quality)

The criteria stated in Table 2.1 were adapted from Law et al., (1998) Critical review for Quantitative Studies. Table 2.2 summarises the overall WoE A judgements based on these criteria. Copies of coded papers can be seen in Appendix 6.

Table 2  
*Weight of Evidence A Criteria*

Dimension 1: Design	Criteria
For a rating of 1 studies should include...	Clearly outlined study design.
For a rating of 2 studies should include...	As above, and appropriate study design choice for current knowledge level.
For a rating of 3 studies should include...	As both above, and consideration for biases that may influence results.
Dimension 2: Sample	Criteria
For a rating of 1 studies should include...	A description of the study sample with participants giving informed consent.
For a rating of 2 studies should include...	As above, with comparisons between groups included to ensure similarity of groups.
For a rating of 3 studies should include...	As both above, and justification of sample size.
Dimension 3: Outcomes	Criteria
For a rating of 1 studies should include...	Clearly outline outcome measures with frequency of assessment recorded.
For a rating of 2 studies should include...	As above, with reliability of measure being assessed.
For a rating of 3 studies should include...	As both above, with validity of measure being assessed.
Dimension 4: Intervention	Criteria
For a rating of 1 studies should include...	Detailed description of the intervention so that it could be replicated in practice.
For a rating of 2 studies should include...	As above, with consideration for contamination bias.

Table 2  
*Weight of Evidence A Criteria*

For a rating of 3 studies should include...	As both above, with consideration for avoiding cointervention.
<b>Dimension 5: Results</b>	<b>Criteria</b>
For a rating of 1 studies should include...	Results reported in terms of significance using an appropriate analysis method.
For a rating of 2 studies should include...	As above, with clinical importance also being discussed.
For a rating of 3 studies should include...	As both above with effect size also being reported.

Table 2.2  
*Summary of Weight of Evidence A Judgements*

Studies	Design	Sample	Outcome	Intervention	Results	WoE A
A Bailey et al. (2012)	2	2	2	1	2	1.8
B Bailey et al. (2015)	2	2	1	1	1	1.4
C Bradley et al. (2003)	2	2	2	1	2	1.8
D Porzig-Drummond et al. (2014)	2	3	2	1	3	2.2
E Porzig-Drummond et al. (2015)	2	3	2	1	3	2.2
F Flaherty and Cooper (2010)	2	1	1	1	1	1.2

Qualitative descriptors (Low: 0-0.9; Medium 1-1.9; High 2-3)

### Appendix 3 – WoE B (Methodological Relevance)

Initial judgements were made using Petticrew and Roberts (2003) criteria, as outlined in Table 3.1. In addition to this an adapted version of the PEDro Scale (Maher et al., 2003) was applied which is outlined in Figure 3.1, with results shown in Table 3.2. A PEDro Score of 0 gave a WoE B score of 1 as it was still classed as a randomised control trial, a PEDro score of 1-2 was given a score of 2, and 3 or above a score of 3.

Table 3.1  
*Petticrew and Roberts, 2003 Criteria*

Study Design	A	B	C	D	E	F	Rationale
Qualitative research, non-experimental evaluations and case-control studies							For answering questions about 'effectiveness' these types of study are deemed to be suitable (Petticrew & Roberts, 2003)
Quasi-experimental designs and cohort studies							
Randomised control studies	x	x	x	x	x	x	

<ol style="list-style-type: none"> <li>1. Explanation of how subjects were randomly allocated to groups</li> <li>2. Allocation was concealed</li> <li>3. The groups were matched at baseline</li> <li>4. There was blinding of all subjects</li> <li>5. There was blinding of all therapists who delivered the intervention</li> <li>6. There was blinding of all assessors who measured the outcomes</li> </ol>	<p>Each satisfied item contributes 1 point to the total PEDro scale.</p> <p>Operational definitions can be seen in Maher et al., 2003.</p>
--	--

Figure 3.1

*Items required to satisfy PEDro Scale.*

Table 3.2  
*Weight of Evidence B Judgement*

Studies	PEDro Criteria Met	PEDro Score	WoE B: Methodological relevance
<b>A</b> Bailey et al. (2012)	3	1	2
<b>B</b> Bailey et al. (2015)	3	1	2
<b>C</b> Bradley et al. (2003)	3	1	2
<b>D</b> Porzig- Drummond et al. (2014)	1, 3, 4	3	3
<b>E</b> Porzig- Drummond et al. (2015)	1, 3	2	2
<b>F</b> Flaherty and Cooper (2010)	No criteria met	0	1

Qualitative descriptors (Low: 0-0.9; Medium 1-1.9; High 2-3)

## Appendix 4 – WoE C (Relevance to Review Question)

Weight of evidence C allows critical consideration of the studies' appropriateness for answering the research question. Therefore, detailed criteria were applied, which can be seen in Table 4.1, before judgements were made, which are presented in Table 4.2.

Table 4.1  
*Criteria for Weigh of Evidence C*

Score	Length of delivery	Supervision	Child Diagnosis	Population
3	Additional training to recommended 6 hours of training received.	Intervention delivery instructed by psychologist.	Child does not have a behaviour disorder diagnosis.	Child protection setting in United Kingdom.
2	Minimum recommendation 6 hours training received.	Intervention delivery instructed by non-psychologist.	Caregiver suspect child may have a behavioural disorder.	Child protection setting not in United Kingdom.
1	Length of delivery varied between participants or not explicitly stated.	Intervention delivery delivered remotely with supervision available via email or not explicitly stated.	No information given.	Not a Child protection setting and not in the United Kingdom.

Table 4.2  
*Summary of Weight of Evidence C Judgements*

Studies	Delivery	Supervision	Diagnosis	Population	WoE C
A Bailey et al. (2012)	2	2	3	1	2
B Bailey et al. (2015)	2	2	3	1	2
C Bradley et al. (2003)	3	2	1	1	1.75
D Porzig-Drummond et al. (2014)	3	3	1	1	2
E Porzig-Drummond et al. (2015)	1	1	1	1	1
F Flaherty and Cooper (2010)	1	1	1	2	1.25

Qualitative descriptors (Low: 0-0.9; Medium 1-1.9; High 2-3)

## Appendix 5 - WoE D Rationale

On WoE A, B and C individuals could score a minimum of one and a maximum of three. Therefore, the maximum score attainable for WoE D is nine and the minimum is three. This means that study scores have a possible range of six. This range was divided into three to determine the Low, Medium and High ratings for WoE D. Cut-offs for each descriptor can be seen in Table 5.1.

Table 5.1  
*Weight of Evidence D Descriptors*

	Minimum Score	Maximum Score
High	7	9
Medium	5	6.99
Low	3	4.99

## Appendix 6 – Coded paper example [A] Bailey et al., (2012)

### *Critical Review Form - Quantitative Studies*

© Law, M., Stewart, D., Pollock, N., Letts, L., Bosch, J., & Westmorland, M., 1998  
McMaster University

#### CITATION:

<p>Bailey, E. L., van der Zwan, R., Phelan, T. W., &amp; Brooks, A. (2012). The 1-2-3 Magic Program: implementation outcomes of an Australian pilot evaluation with school-aged children. <i>Child &amp; Family Behavior Therapy</i>, 34(1), 53-69.</p>

#### Comments

<p><b>STUDY PURPOSE:</b> Was the purpose stated clearly? <input type="radio"/> Yes <input type="radio"/> No</p>	<p>Outline the purpose of the study. How does the study apply to occupational therapy and/or your research question?</p>
<p><b>LITERATURE:</b> Was relevant background literature reviewed? <input type="radio"/> Yes <input type="radio"/> No</p>	<p>Describe the justification of the need for this study.</p>
<p><b>DESIGN:</b> <input checked="" type="radio"/> randomized (RCT) <input type="radio"/> cohort <input type="radio"/> single case design <input type="radio"/> before and after <input type="radio"/> case-control <input type="radio"/> cross-sectional <input type="radio"/> case study</p>	<p>Describe the study design. Was the design appropriate for the study question? (e.g., for knowledge level about this issue, outcomes, ethical issues, etc.)</p> <p>The use of a randomized control trial was used appropriately despite the small study size, as there has been previous research in this field. Participants were assigned to a treatment group or a wait-list control meaning they were all exposed to the intervention over a 3 month period. This is a more ethical way of managing a control group whilst ensuring they still have access to the intervention which is believed by eh researchers to be beneficial.</p> <p>Specify any biases that may have been operating and the direction of their influence on the results.</p> <p>Volunteer bias: All parents were self-selected and were all female. Both may have an impact on how a child's behaviour is perceived. As the measures are all self-report it is this could have an effect, in either direction, on the effectiveness of this study.</p>

**Comments**

<p><b>SAMPLE:</b> N = 9</p> <p>Was the sample described in detail?  <input checked="" type="radio"/> Yes  <input type="radio"/> No</p> <p>Was sample size justified?  <input type="radio"/> Yes  <input checked="" type="radio"/> No  <input type="radio"/> N/A</p>	<p>Sampling (who; characteristics; how many; how was sampling done?) If more than one group, was there similarity between the groups?</p> <p>Demographic data was presented as part of the results section of this study. This demographic data represented the 9 self-selected female participants who took part. Statistical significance was not determined for differences seen between the control and treatment group.</p> <p>Describe ethics procedures. Was informed consent obtained?          Informed consent was given after participants had responded to the initial newspaper advert, however there is no description of how this was obtained. There is no reference to the ethical governing body behind this study.</p>								
<p><b>OUTCOMES:</b></p> <p>Were the outcome measures reliable?  <input checked="" type="radio"/> Yes  <input type="radio"/> No  <input type="radio"/> Not addressed</p> <p>Were the outcome measures valid?  <input type="radio"/> Yes  <input type="radio"/> No  <input checked="" type="radio"/> Not addressed</p>	<p>Specify the frequency of outcome measurement (i.e., pre, post, follow-up)          Baseline assessments were obtained prior to intervention delivery and were then repeated 6 weeks after the intervention had been delivered.</p> <table border="1" data-bbox="544 779 1492 1137"> <thead> <tr> <th data-bbox="544 779 1050 808">Outcome areas (e.g., self-care, productivity, leisure).</th> <th data-bbox="1050 779 1492 808">List measures used.</th> </tr> </thead> <tbody> <tr> <td data-bbox="544 808 1050 869">Child behaviour</td> <td data-bbox="1050 808 1492 869">Eyberg Child Behaviour Inventory (ECBI) (Eyberg &amp; Pincus, 1999)</td> </tr> <tr> <td data-bbox="544 869 1050 929">Parental attitude</td> <td data-bbox="1050 869 1492 929">Parenting sense of competence scale (PSOC) (Johnston and Mash, 1989)</td> </tr> <tr> <td data-bbox="544 929 1050 1137">Demographics Questionnaire</td> <td data-bbox="1050 929 1492 1137">Parenting information questionnaire (PIQ) (unpublished)</td> </tr> </tbody> </table>	Outcome areas (e.g., self-care, productivity, leisure).	List measures used.	Child behaviour	Eyberg Child Behaviour Inventory (ECBI) (Eyberg & Pincus, 1999)	Parental attitude	Parenting sense of competence scale (PSOC) (Johnston and Mash, 1989)	Demographics Questionnaire	Parenting information questionnaire (PIQ) (unpublished)
Outcome areas (e.g., self-care, productivity, leisure).	List measures used.								
Child behaviour	Eyberg Child Behaviour Inventory (ECBI) (Eyberg & Pincus, 1999)								
Parental attitude	Parenting sense of competence scale (PSOC) (Johnston and Mash, 1989)								
Demographics Questionnaire	Parenting information questionnaire (PIQ) (unpublished)								
<p><b>INTERVENTION:</b></p> <p>Intervention was described in detail?  <input checked="" type="radio"/> Yes  <input type="radio"/> No  <input type="radio"/> Not addressed</p> <p>Contamination was avoided?  <input type="radio"/> Yes  <input type="radio"/> No  <input checked="" type="radio"/> Not addressed  <input type="radio"/> N/A</p> <p>Cointervention was avoided?  <input type="radio"/> Yes  <input type="radio"/> No  <input checked="" type="radio"/> Not addressed  <input type="radio"/> N/A</p>	<p>Provide a short description of the intervention (focus, who delivered it, how often, setting). Could the intervention be replicated in occupational therapy practice?</p> <p>Intervention was practitioner-led during 3 x 2-hour sessions taking place over a 2-day period. A combination of '1-2-3 Magic' DVDs, PowerPoint presentations, handouts, small-group activities and active skills training were used throughout the sessions.          Practitioners can come from a number of backgrounds including teachers, child-care workers and psychologists, and become commercially trained via a self-directed Leader Guide.</p>								

**Comments**

<p><b>RESULTS:</b></p> <p>Results were reported in terms of statistical significance?</p> <p><input checked="" type="radio"/> Yes</p> <p><input type="radio"/> No</p> <p><input type="radio"/> N/A</p> <p><input type="radio"/> Not addressed</p> <p>Not including effect sizes</p> <p>Were the analysis method(s) appropriate?</p> <p><input checked="" type="radio"/> Yes</p> <p><input type="radio"/> No</p> <p><input type="radio"/> Not addressed</p> <p>Clinical importance was reported?</p> <p><input checked="" type="radio"/> Yes</p> <p><input type="radio"/> No</p> <p><input type="radio"/> Not addressed</p>	<p>What were the results? Were they statistically significant (i.e., <math>p &lt; 0.05</math>)? If not statistically significant, was study big enough to show an important difference if it should occur? If there were multiple outcomes, was that taken into account for the statistical analysis?</p> <p>There was a significant decrease of score on intensity and problem scales of behaviour (<math>p &lt; 0.05</math> for both), for individuals in the treatment group. Whilst the slight improvement of behaviour in the wait-control group was not significant (however significance levels were not reported).</p> <p>There was a significant increase in scores of parental attitude towards children in the treatment group (<math>p &lt; 0.05</math> for all 3 scales), and although slight differences were seen in the wait-list control these findings were not significant (again no p scores were reported).</p> <p>These changes were seen to be maintained at follow up, with no significance in change of magnitude.</p> <p>These results both come from t-tests that compare one variable before and after intervention or control. A more meaningful way of comparing multiple outcomes based on different variables could have been achieved with an ANOVA. There is also no reporting of overall effect size.</p> <p>The non-significant findings for the wait control group could be due to the small, unjustified sample size.</p> <p>What was the clinical importance of the results? Were differences between groups clinically meaningful? (if applicable)</p> <p>All children's behaviours were considered to be within the clinical range before the intervention was delivered, and following the delivery of the intervention both the intensity and problem scale scores were reduced to the non-clinical range. However as there was no significant change show in the control group, it is assumed that they remain within the clinical range for intensity and problem scales of behaviours.</p>
<p>Drop-outs were reported?</p> <p><input checked="" type="radio"/> Yes</p> <p><input type="radio"/> No</p>	<p>Did any participants drop out from the study? Why? (Were reasons given and were drop-outs handled appropriately?)</p> <p>There were five dropouts in this study, all of which have had their reasons for attrition recorded.</p>
<p><b>CONCLUSIONS AND CLINICAL IMPLICATIONS:</b></p> <p>Conclusions were appropriate given study methods and results</p> <p><input type="radio"/> Yes</p> <p><input type="radio"/> No</p>	<p>What did the study conclude? What are the implications of these results for occupational therapy practice? What were the main limitations or biases in the study?</p>