



Title: An evidence review of the impact Family Group Conferencing (FGC) and Restorative Practices (RP) have on positive outcomes for children and families - **Appendices**

Date: 26th July 2016

Appendix A. The REA method

This report presents the results of a Rapid Evidence Assessment (REA). An REA is a tool for establishing the quality and quantity of available empirical research evidence on a policy issue, as comprehensively as possible within the constraints of a given timetable. The REA was commissioned as part of the evaluation of the Leeds Family Valued programme, funded by the DfE Social Care Innovation Programme and led by ICF.

REAs use systematic review methods to search and critically appraise existing research. The strength of the method lies in following a clear set of procedures and recording precisely what has been done at each step, so that the process can be repeated if necessary. REAs are characterised by rigorous and explicit methods, but provide a quick synthesis of the available evidence by shortening the traditional systematic review process. They can constrain particular aspects of the process by, for example:

- limiting the breadth of the research question;
- using less developed search strings rather than extensive search of all variants;
- using “grey” and print sources but less exhaustively than systematic reviews;
- establishing good “inter-rater” reliability for quality assessment by having two people rate a sample of relevant papers. Once reliability has been established, time and resources are saved by using one ‘rater’ to assess the majority of papers.

This review has been conducted according to guidelines developed and written by Government Social Research (GSR) and the Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre), part of the Social Science Research Unit at the Institute of Education, University of London¹.

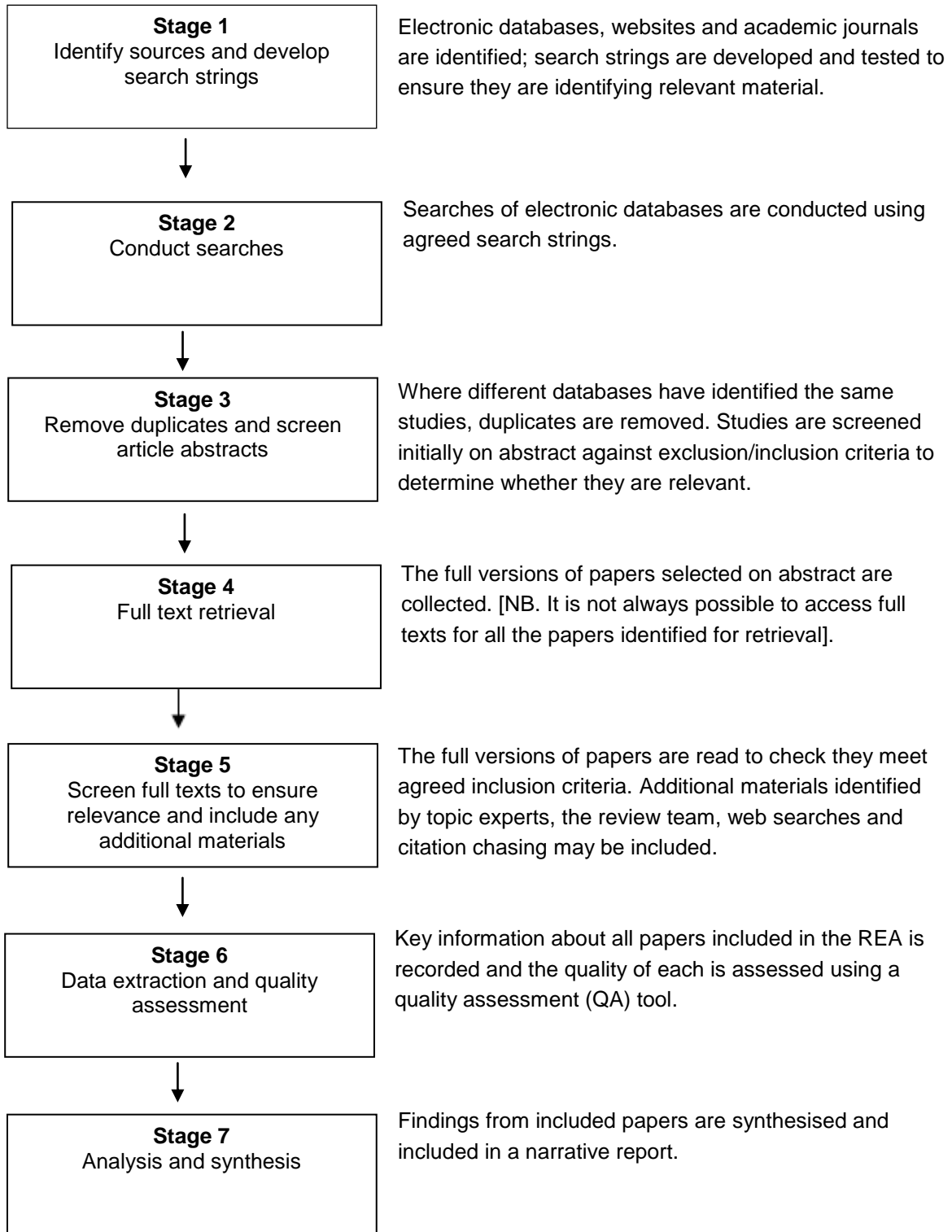
It is worth noting, however, that restricted resources can create limitations for REAs compared with full systematic reviews:

- Time constraints mean REAs (a) may miss some literature not catalogued on the key electronic databases, and (b) the majority of quality ratings are conducted by one assessor, with a second assessor rating only a small subset;
- Some of the studies included are inevitably of limited methodological quality. As a consequence, results should be generalised with caution;
- Time does usually not allow for REAs to involve “pearl growing”, ie looking through the reference lists of selected articles for other potentially important sources that electronic searches may have missed; and
- All review methods, including REAs, risk generating inconclusive findings that provide a weak answer to the original question if there are too few good-quality studies to address it. In such cases, the tight REA timescales mean that there may be no opportunity to go back and reformulate the question or inclusion criteria.

¹ <http://webarchive.nationalarchives.gov.uk/20140305122816/http://www.civilservice.gov.uk/networks/gsr/resources-and-guidance/rapid-evidence-assessment>

Figure 2 summarises the stages of the REA process. Full details of the REA methods, including copies of the quality assessment tools we have used, are in the appendices to this report.

Figure 2: Summary of REA methodology.



A.1 Assessing the strength of a body of evidence

The past 20 years have seen a real growth in what has become known as Evidence-Based Practice (EBP). Stakeholders have recognised the benefits of developing EBP in areas such as public health and social policy. To quote *The Magenta Book*:

“Good evaluation, and the reliable evidence it can generate, provides direct benefits in terms of policy performance and effectiveness, but is also fundamental to the principles of good government, supports democratic accountability and is key to achieving appropriate returns from taxpayers’ resources. A good evaluation is therefore a normal and natural part of policy making and effective government and is a powerful tool available to the policy maker.” (p.12)²

Evidence reviews such as an REA, focusing on empirical impact studies, are a critical element in developing EBP; they are used to summarise the main characteristics of a body of evidence in relation to a specific issue. Guidance on how to assess the strength of a body of evidence typically highlights four important characteristics (Lohr, 2004; DfID, 2013):

- (a) the **quality** of individual articles or papers that make up the body of evidence;
- (b) the **quantity** (number) of papers that make up the body of evidence;
- (c) the **consistency** of the findings produced by the studies making up the body of evidence;
- and
- (d) the **context** in which the available evidence has been collected.

A.1.1 The quality of individual studies that make up the body of evidence

Based on established evaluative methods (Lohr, 2004), we used two quality-assessment systems, one for primary research studies, and a second for evidence reviews, to assess studies included in the review. As noted above, an essential element of a review of this type is to provide a guide to the credibility of each study included.

Quality assessments of primary research studies

We assessed the quality of primary research studies on seven criteria: rationale for overall research strategy, study design, sampling strategy, data collection procedures, data analysis, interpretation and reporting of results, and credibility of conclusions. Where primary studies tested the impact of specific interventions, in addition to the seven criteria listed above, we rated the design of the intervention study using the Maryland Scientific Methods Scale (SMS)³. Not all primary studies test interventions (e.g. some may report survey findings), therefore not all primary

² https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/220542/magenta_book_combined.pdf

³ Sherman, L. Gottfredson, D. MacKenzie, D. Eck, J. Reuter, P. Bushway, S. (1997) *Preventing Crime: What Works, What Doesn't, What's Promising* Washington: US Department of Justice.

studies were rated on the SMS. Details of the quality-assessment system for primary studies and quality scores for papers assessed can be found in Appendix C, along with a description of the SMS scoring system.

Quality assessments of reviews

For reviews, we used eight criteria: review method, search strategy, data collection (sift), quality appraisal, data analysis (quantitative), qualitative synthesis, interpretation and reporting of results, and credibility of conclusions. Details of the quality-assessment system we used for reviews can be found in Appendix C of this report, along with quality scores for all the reviews included.

A. 1.2 The quantity of papers that make up the body of evidence

One of the key strengths of empirical research that is the focus of an REA is the capacity to replicate or repeat investigations to see if the results are consistent. That is why it is so important that research papers provide enough detail of how an investigation was conducted to enable someone else to repeat it. The more often a finding is replicated, the more confident we can be that the effect is real, rather than a product of the study's design and implementation; the more studies done to test a particular theory or intervention, the stronger the body of evidence. However, there is no rule of thumb for how many studies constitute an adequate body of evidence. That often depends on the research question being investigated; a more complex question will require more studies before the researchers can be confident that the evidence base is strong. Certainly, where only one or two studies have been done, even if they were well-designed, it is reasonable to conclude that the body of evidence is small or weak. Based on recommendations, we take a case-by-case approach.⁴ For each review we undertake we categorise the size of the evidence base as small, medium or large, and specify the number of studies associated with each category. Typically, 'small' might be where the review has identified five studies or fewer, 'medium' between six and ten, and 'large' if eleven or more studies were found.

A. 1.3 The consistency of the findings produced by the studies making up the body of evidence

A strong body of evidence is usually defined as one where a large number of studies all report the same or similar findings when a specific intervention is delivered to a particular group of end users. However, social interventions like RP are typically complex. As a result, it is possible to have a large number of studies that, because they have tested slightly different interventions in different social contexts, do not provide entirely consistent findings. Using a review to synthesise the findings from multiple studies helps to establish the degree of consistency in a body of evidence by exploring the impact of these similarities and differences. That we found fewer empirical impact studies of RP than of FGC is likely to be, at least in part, because RP is more of an umbrella term rather than a specific intervention.

⁴ Department for International Development (2013). Assessing the strength of evidence: DfID practice paper. www.gov.uk/government/publications/how-to-note-assessing-the-strength-of-evidence . Last accessed March 10th 2014

A. 1.4 The context in which the available evidence has been collected

A review needs to acknowledge the context in which the evidence cited has been produced. It is important to have a good understanding of how well evidence collected in one particular context can be generalised to another. In social policy research, country of origin is often, although by no means always, relevant. With regard to RP, critical elements of social context may include details of the wider landscape of social services within which these approaches are being used. Depending on the level of detail reported in individual papers, it may not always be possible to take such variations into account.

A. 1.5 Summary

To summarise, the strength of a body of evidence depends on the quantity of research that has been conducted, the quality of that research, the context in which the research was done, and consistency of findings across papers and articles uncovered by a search of appropriate sources. The rest of Section 4.0 describes the body of evidence in terms of three of those criteria: quantity, quality and context. We will explore the matter of consistency in research findings with regard to particular interventions and issues in Section 6.0.

A.2 Quantity of research available

Our review of evidence concerning the impact of restorative practices (including FGC and FGDM) on positive outcomes for children and families was limited to studies that had the following characteristics:

- i. reported in the English language;
- ii. reported data as part of an empirical investigation, or reviewed empirical studies that reported data;
- iii. included children and young people aged from 0 – 24 years in their sample;
- iv. included investigation of the impact of restorative practices (including FGC and FGDM) on at least one of the following: (i) positive outcomes for children; (ii) removal of children from families; (iii) domestic violence in families; (iv) on numbers of children entering the criminal justice system; (v) reducing welfare concerns early in the aetiology of a family problem; (vi) improved support for children, young people and families;
- v. published after 1999 in the case of empirical papers, and 2005 in the case of reviews⁵.

Table 1 summarises the key search terms we used and the number of hits returned from each of the databases we searched.

⁵ We included one review published prior to that date because it looked specifically at the impact of legislation in New Zealand, the first country to legislate for the use of Family Group Conferences.

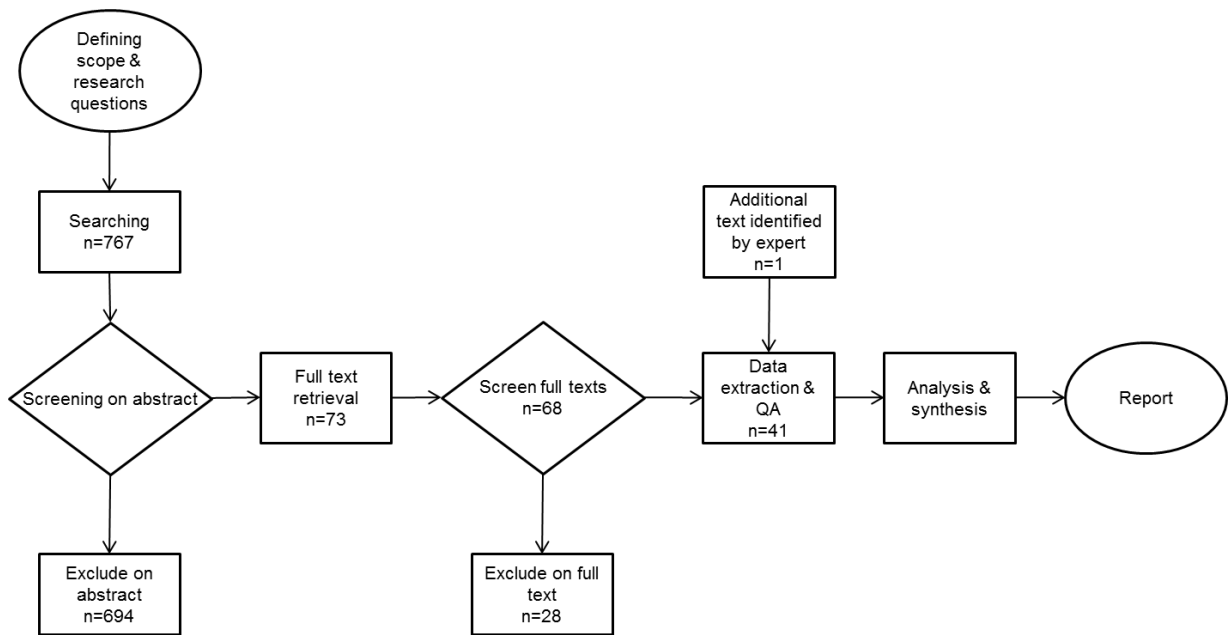
Full details of our search results and the inclusion criteria we used are provided in Appendices A.2 and A.3.

Table 1
Summary search terms and hits returned by database searched

Lead search term	Database	Hits
Restorative practice	Social Policy & Practice	57
	Applied Social Sciences Index & Abstracts	27
	Social Services Abstracts	28
	Sociological Abstracts	44
	Subtotal	156
Family Group Conference	Social Policy & Practice	216
	Web of Science	58
	Social Services Abstracts	63
	Applied Social Sciences Index & Abstracts	57
	Sociological Abstracts	30
Subtotal	424	
Family Centred Approach	Social Policy & Practice	81
	Applied Social Sciences Index & Abstracts	7
	Web of Science	22
	Social Services Abstracts	7
	Sociological Abstracts	7
Subtotal	124	
Family Group Decision Making	Social Policy & Practice	31
	Applied Social Sciences Index & Abstracts	10
	Social Services Abstracts	21
	Sociological Abstracts	1
Subtotal	63	
	TOTAL	767

The flow diagram below shows the numbers of studies identified at each stage of the REA.

Figure 3: Rapid Evidence Assessment (REA) workflow: review of evidence concerning the impact of restorative practices, specifically Family Group Conferencing (FGC) and Family Group Decision Making (FGDM), on outcomes for children and families.



Of the 73 references selected for full text retrieval, we were able to obtain 68. When the full texts were reviewed, a further 28 papers were excluded because they did not meet the inclusion criteria, leaving 40 sources for inclusion in the review. Peer reviewers of an earlier draft identified a further review paper. Of those 41 papers in total, eight were reviews of research, and 33 described primary research studies. The inclusion and exclusion criteria are given in Appendix B.3.

A. 2.1 Quality of individual research studies and reviews

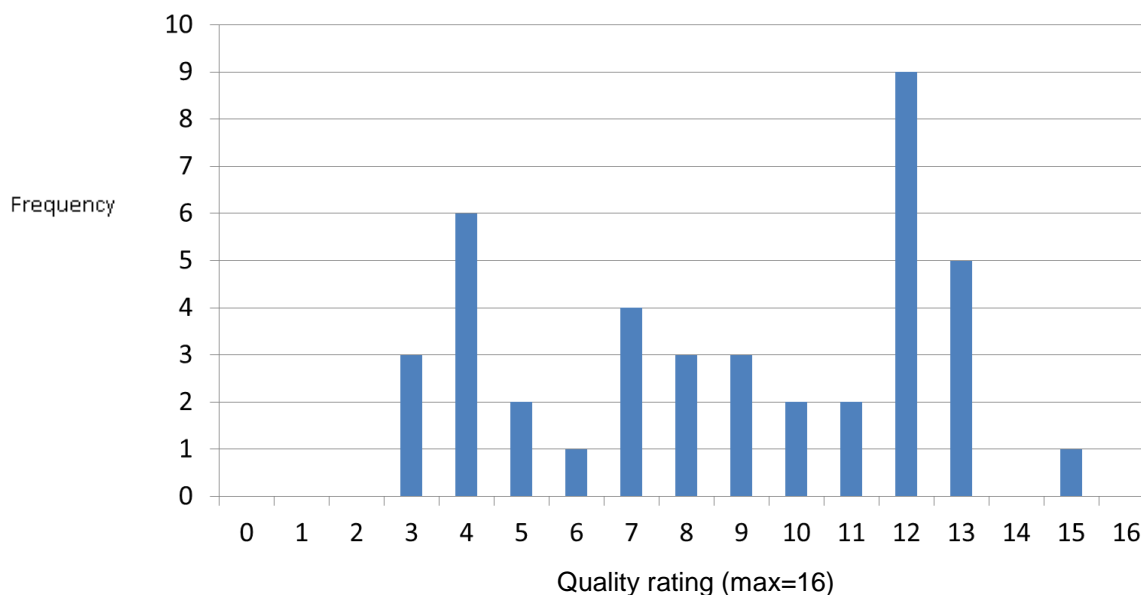
We assessed primary studies on seven quality criteria, each marked on a scale of 0-2, giving a maximum possible quality score of 14. The full table including details of component quality scores is included in Appendix C.

We assessed each review on eight criteria and, as before, each criterion was marked on a scale of 0-2, giving a possible total score of 16. A full version of the table providing the breakdown of quality scores appears in Appendix D.

Figure 4 (below) shows the distribution of the quality ratings across all 41 papers we included in the review; the majority report primary research rather than reviews. The primary research

studies we included were generally of a good standard. While too few good-quality studies have been published to enable us to reach any definitive conclusion regarding the impact of restorative approaches on child and family outcomes, that position is changing. For example, two recently published papers describe randomised controlled trials (Jeong *et al.*, 2012; Malmberg-Heimonen & Johansen, 2014).

Figure 4: Distribution of quality scores across papers included in the review (n=41)



However, our REA searches found relatively few recent evidence reviews. Of those we did find, all but one came from either the UK or the US. We found one systematic review, but no others that included robust evaluation of the included studies. It is therefore reasonable to conclude that, with the exception of the one Campbell Collaboration systematic review, the quality of existing evidence reviews is not high according to accepted criteria.

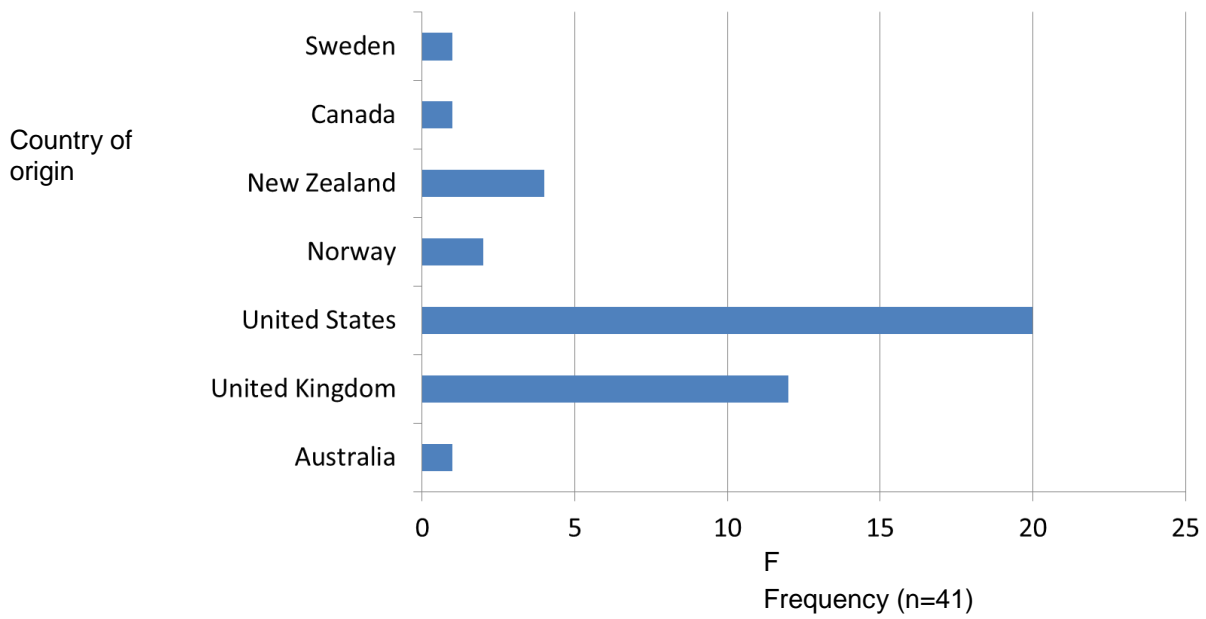
3.1 Context – international comparisons

Figure 5 (below) shows that most empirical research in this area comes from the US and the UK. Of the 41 papers included in the REA, 12 originated in the UK. As a consequence, we need to consider carefully how far research conducted in one global region might be relevant to policy and practice in another. The research studies we have reviewed show quite clearly how the delivery of restorative practices, including FGCs and FGDMs, varies considerably. As one review (Barnsdale & Walker 2007) remarked:

“The original FGC model has been adapted many times so that it can be applied within different jurisdictions in relation to a variety of issues at different points in relevant child welfare/juvenile justice systems..” (p. 2)

To that end, we begin our narrative review by identifying, as far as possible, common elements of restorative approaches that will help establish the relevance of international research findings to practice in the UK.

Figure 5: Countries of origin of papers included in the review



Appendix B: REA results

B.1 Databases searched

Our search partners from the King's College Information Retrieval Unit have searched the following databases:

Applied Social Sciences Index and Abstracts (ASSIA)

Applied Social Sciences Index and Abstracts (ASSIA) is an indexing and abstracting tool covering health, social services, psychology, sociology, economics, politics, race relations and education. It provides a comprehensive source of social science and health information that is updated monthly. ASSIA currently contains over 375,000 records from over 500 journals published in 16 different countries, including the UK and US.

Social Policy and Practice

Social Policy and Practice (SPP) is a specialist bibliographic database covering the social care, education, health and support sectors. It combines the unique databases of four major UK organisations active in social policy:

Ageinfo – Centre for Policy on Ageing;
ChildData – National Children's Bureau;
Planex – IDOX Information Service;
Social Care Online – Social care Institute for Excellence.

SPP contains more than 300,000 bibliographic records and abstracts dating from 1981; more than 24,000 new records are added each year. It includes documents and publications from national and regional government departments, local authorities, public agencies, third-sector organisations, consultancy and professional bodies, academic think tanks and research departments.

Social Services Abstracts

Social Services Abstracts provides bibliographic coverage of current research focused on social work, human services, and related areas, including social welfare, social policy, and community development. The database abstracts and indexes more than 1,300 serials publications and includes abstracts of journal articles and dissertations, and citations to book reviews. Major areas of coverage include community and mental health services, family and social welfare, poverty and homelessness, professional issues in social work, social and health policy, social work practice, and violence, abuse, and neglect. The database is updated monthly with approximately 5,500 records added per year.

Sociological Abstracts

Sociological Abstracts provides abstracts from the international literature in sociology and related disciplines in the social and behavioural sciences. It covers journal articles and citations to book reviews drawn from over 1,800 journals, as well as providing abstracts of books, book chapters, dissertations, and conference papers. It draws abstracts from a variety of sources including

journal articles, conference papers, books, dissertations, and conference papers, plus citations to important book reviews related to the social sciences. A back file begins in 1952 with records published by the then print version of Sociological Abstracts; 40 per cent of the provided content is published outside North America. Areas of coverage include culture and social structure, economic development, evaluation research, family and social welfare, health, medicine and law, methodology and research technology, and substance abuse and addiction. The database is updated monthly, with approximately 30,000 records added per year.

Web of Science

Web of Science (previously known as Web of Knowledge) is an online subscription-based scientific citation indexing service that gives access to multiple databases referencing cross-disciplinary research. Web of Science has indexing coverage from the year 1900 to the present. The multidisciplinary coverage of the Web of Science encompasses over 50,000 scholarly books, 12,000 journals and 160,000 conference proceedings. The coverage includes the sciences, social sciences, arts, and humanities, and goes across disciplines. As of September 3, 2014 the total file count of the Web of Science was 90 million records, which included over a billion cited references. The citation service indexes around 65 million items per year, earning it the description of the largest accessible citation database.

B.2. Search terms and hits by database

We used the following search terms:

Database	Search terms	No. of hits
Social Policy and Practice 2005-2015	1 family group confer*. (414) 2 limit 1 to yr="2005 -Current" (216)	216
Web of Science 2005-2015	1 FGC (58)	58
Social Services Abstracts 1999-2015	1 FGC (63)	63
Applied Social Sciences Index and Abstracts (ASSIA) 1999-2015	1 FGC (57)	57
Sociological Abstracts 1999-2015	1 FGC (30)	30
Social Policy and Practice 2007 - 2014	1 (restorative and (practice* or approach*)). (349) 2 limit 1 to yr="2005 -Current" (268) 3 justice (12044) 4 limit 3 to yr="2005 -Current" (6337) 5 2 not 4 (33) 6 (restorative and (practice* or approach* or supervision or care or meeting* or conferenc*)).mp. [mp=abstract, title, publication type, heading word, accession number] (442) 7 limit 6 to yr="2005 -Current" (339) 8 7 not 4 (57) 9 (restorative and (practice* or approach* or supervision or care or meeting* or domestic or conferenc*)). (447) 10 limit 9 to yr="2005 -Current" (342) 11 10 not 4 (57)	57
Social Policy and Practice 2005-2015	1 family centred approach*.mp. [mp=abstract, title, publication type, heading word, accession number] (131) 2 limit 1 to yr="2005 -Current" (83) 3 (restorative or justice).mp. [mp=abstract, title, publication type, heading word, accession number] (12134) 4 limit 3 to yr="2005 -Current" (6406)	81

Database	Search terms	No. of hits
	5 2 not 4 (81)	
Social Policy and Practice 2005-2015	1 FGDM (31)	31
Web of Science 2005-2015	1 FCA (22)	22
Applied Social Sciences Index and Abstracts (ASSIA) 2005-2015	1 ((restorative AND (practice* OR approach* OR supervision OR care OR meeting* OR domestic OR conferenc*)) NOT (justice OR crim*)) AND la.exact("English") AND pd(2005-2015) (87) 2 ((restorative AND (practice* OR approach* OR supervision OR care OR meeting* OR domestic OR conferenc*)) NOT (justice OR crim* OR health OR korea* OR india* OR afric* OR medic* OR clinic*)) AND la.exact("English") AND pd(2005-2015) (27)	27
Applied Social Sciences Index and Abstracts (ASSIA) 2005-2015	1 (("family centered approach" OR "family centred approach") NOT (justice OR crim* OR health OR korea* OR india* OR afric* OR medic* OR clinic*)) AND la.exact("English") AND pd(2005-2015) (7)	7
Applied Social Sciences Index and Abstracts (ASSIA) 2005-2015	1 ("family group decision making" NOT (justice OR crim* OR health OR korea* OR india* OR afric* OR medic* OR clinic*)) AND la.exact("English") AND pd(2005-2015) (10)	10
Social Services Abstracts 2005-2015	1 ((restorative AND (practice* OR approach* OR supervision OR care OR meeting* OR domestic OR conferenc*)) NOT (justice OR crim*)) AND la.exact("English") AND pd(2005-2015) (28)	28
Social Services Abstracts 2005-2015	1 (((("Family centred approach" OR "family centered approach") NOT (justice OR crim* OR health OR medic*)) AND rtype.exact("Review" OR "Journal Article"))) AND la.exact("English") AND pd(2005-2015) (7)	7
Social Services Abstracts	1 ("family group decision making" NOT (justice OR crim* OR health OR korea* OR india* OR afric* OR medic* OR clinic*)) AND la.exact("English") AND pd(2005-2015) (21)	21

Database	Search terms	No. of hits
2005-2015		
Sociological Abstracts 2005-2015	1 ("family centered approach" OR "family centred approach") NOT (justice OR crim* OR health OR korea* OR india* OR afric* OR medic* OR clinic*) AND la.exact("English") AND pd(2005-2015) (7)	7
Sociological Abstracts 2005-2015	1 ((restorative AND (practice* OR approach* OR supervision OR care OR meeting* OR domestic OR conferenc*)) NOT (justice OR crim*)) AND la.exact("English") AND pd(2005-2015) (44)	44
Sociological Abstracts 2005-2015	1 ("family group decision making" NOT (justice OR crim* OR health OR korea* OR india* OR afric* OR medic* OR clinic*)) AND la.exact("English") AND pd(2005-2015) (1)	1
	TOTAL	767

B.3. Inclusion and exclusion criteria

Inclusion Criteria

Criteria	Description	Example
Impact of restorative practices (including FGC and FGDM) on positive outcomes for children	Studies and reviews that investigate direct links between restorative practices and outcomes for positive outcomes children.	Johansen, S. (2014). Psycho-social processes and outcomes of family group conferences for long-term social assistance recipients. <i>British Journal of Social Work</i> , 44 (1), 145-162.
Impact of restorative practices (including FGC and FGDM) on removal of children from families	Studies and reviews that investigate the direct links between restorative practices and removal of children from families.	Pennell, J., Edwards, M., & Burford, G. (2010). Expedited family group engagement and child permanency. <i>Children and Youth Services Review</i> , 32 (7), 1012-1019.
Impact of restorative practices (including FGC and FGDM) on domestic violence in families	Studies and reviews that investigate the direct links between restorative practices and the reduction of domestic violence in families	Weigensberg, E. C., Barth R., Guo S. (2009). Family group decision making: a propensity score analysis to evaluate child and family services at baseline and after 36-months. <i>Children and Youth Services Review</i> , 31 (3), 383-390.
Impact of restorative practices (including FGC and FGDM) on numbers of children entering the criminal justice system	Studies and reviews that investigate the direct links between restorative practices and numbers of children entering the criminal justice system	Mutter R., Shemmings, D., Dugmore, P., & Hyare, M. (2008). Family group conferences in youth justice. <i>Health and Social Care in the Community</i> , 16 (3), 262-270.
Impact of restorative practices (including FGC and FGDM) on reducing welfare concerns early in the aetiology of a family problem	Studies and reviews that investigate the direct links between restorative practices and a reduction in welfare concerns	Pountney, K. (2005). <i>Education and early intervention projects: an evaluation of the first year</i> . Essex County Council. Family Group Conference Service
Impact of restorative practices (including FGC and FGDM) on improved support for children, young people and families	Studies and reviews that investigate the links between restorative practices and improved support for children at risk	Pennell, J. (2006). Restorative practices and child welfare: toward an inclusive civil society. <i>Journal of Social Issues</i> , 62 (2), 259-279.
Age range 0 – 24 years	Studies and reviews that investigate the links between restorative practices and positive outcomes for children, particularly in relation to the age range 0 – 24.	Sundell, K., & Vinnerljung, B. (2004). Outcomes of Family Group Conferencing in Sweden. A 3-Year Follow-Up. <i>Child Abuse & Neglect</i> , 28 (3), 267-287.

Exclusion Criteria

Criteria	Description	Example
General think pieces on the philosophy of RP	Discussions of the principles that underpin the RP approach	Connolly, M., Masson, J. (2014). Private and public voices: does family group conferencing privilege the voice of children and families in child welfare? <i>Journal of Social Welfare and Family Law</i> . 36(4),
Workings of the family court system	Implications for family social work of the current court system relating to family law	Doughty, J. (2014) Care proceedings—is there a better way? <i>Child and Family Law Quarterly</i> , 26(2), 113-131.
Theoretical underpinnings of FGC	Examination of the social theories that have been posited to provide a rationale for FGC and restorative approaches	Burns, G., & Fruchtel, F. (2014). Family group conference: a bridge between lifeworld and system. <i>British Journal of Social Work</i> , 44(5), 1147-1161.
FGC in mental health	Papers that examine the use of FGC in mental health settings	De Jong, G., & Schout, G. (2013). Researching the applicability of family group conferencing in public mental health care. <i>British Journal of Social Work</i> , 43(4), 796-802.
FGC in the context of adult safeguarding	Studies and reviews that investigate adults	Hobbs, A., & Alonzi, A. (2013). Mediation and family group conferences in adult safeguarding. <i>Journal of Adult Protection</i> , 15(2)
FGC in the context of physical health	Papers that examine the use of FGC in health settings	Vis, S., A., Strandbu, A., Holtan, A., Thomas, N. (2011). Participation and health: a research review of child participation in planning and decision-making. <i>Child and family social work</i> . 16(3), 325-335.
Ethics of restorative approaches	Ethical consideration of FGC and restorative practices	Evans, C.A. (2009). Ethical implications of child welfare policies in England and Wales on child participation rights. <i>Ethics and Social Welfare</i> , 3(1), 95-101.



Appendix C – quality scores for primary research studies

Quality Appraisal criteria		Rating 0-2
RESEARCH RATIONALE	*Convincing rationale for overall research strategy and how it was designed to meet study aims/ research questions, including comprehensive review of previous research and justification for collecting new primary data.	
RESEARCH DESIGN	*Good discussion of main features of research design and strengths and weaknesses of data sources. *Research design shows robustness (reliable and replicable) and validity. *Implications of limitations taken into consideration in the analysis and findings. *ETHICS - confidentiality, anonymity, data protection, instructions to participants etc, impartiality	
SAMPLING	(a) *Does the study describe locations and populations(s) of interest and how and why chosen (eg, typical or extreme case or diverse constituencies etc) to allow comparisons to be made? (b) *Was the sampling strategy appropriate to research question, e.g. purposive vs random; is size large enough for generalisability if required? (c) *Is the achieved sample representative of the population of interest? Is there information about the response rate?	
DATA COLLECTION	*Detailed description of data and collection methods used, explaining any limitations and methods to maximise inclusion/ limit bias *Reliability - was there pilot testing of tools/methods; did more than one person collect data?	
DATA ANALYSIS	*Explicit and appropriate analytic procedure for processing raw data into results / themes that could be repeated with a similar methodology *Reliability - was there triangulation of data analysis (e.g. multiple scorers or coders)	
INTERPRETATION AND REPORTING OF RESULTS	*Study reports findings on all variables or concepts investigated and includes discussion/ mention of any negative cases and outliers and confounding/moderating variables. *Discussion of mechanisms through which effects happen, with examples from the data *Limitations - discusses importance of study's context and biases/flaws in design	
CREDIBILITY OF CONCLUSIONS	*Conclusions presented are supported by study findings and previous research and theory (where appropriate). *Evidence of openness to new/ alternative ways of viewing subject/ theories/ assumptions *An attempt is made to quantify/ explain the strength or value of the findings, if appropriate?	

Quality assessment summary for primary studies

All criteria marked on a scale of 0 – 2

	Reference	Research rationale	Research design	Sampling	Data collection	Data analysis	Interpretation and reporting of results	Credibility of conclusions	Overall score	Country of study	Study methods	Maryland score if appropriate
4	Baffour ,T. D.(2006).Ethnic and gender differences in offending patterns: examining family group conferencing interventions among at-risk adolescents. <i>Child and Adolescent Social Work Journal</i> , 23 (5-6), 557-578.	2	2	1	1	2	2	2	12	US	Secondary analysis of data collected by police department FGC project.	4
6	Bell, M., & Wilson, K. (2006). Children's views of family group conferences. <i>British Journal of Social Work</i> , 36 (4), 671-681.	2	1	1	0	0	2	2	8	UK	Non-experimental (no comparison group) study of families views of FGC process	n/a
7	Berzin, S. C., Cohen, E., Thomas, K., Dawson, W. C. (2008). Does family group decision making affect child welfare outcomes? Findings from a randomized control study. <i>Child Welfare</i> , 87 (4), 35-54.	2	2	1	2	2	2	2	13	US	Reports two FGDM (Family Group Decision Making) projects evaluated by RCT	5
8	Berzin, S.C. (2006). Using sibling data to understand the impact of family group decision-making on child welfare outcomes. <i>Children and Youth Services Review</i> . 28 (12), 1449-1458.	2	2	1	2	2	2	2	13	US	As above, but with sibling data added	5
9	Berzin, S.C., Thomas, K.L., & Cohen E. (2007). Assessing model fidelity in two family group decision-making programs: is this child welfare intervention being implemented as intended? <i>Journal of Social Service Research</i> , 34 (2), 55-71.	2	2	2	2	2	2	2	13	US	Reports two FGDM projects evaluated by RCT (as above but focused on impact of delivery fidelity)	5
17	Crampton, D. (2006). When do social workers and family members try family group decision making? A process evaluation. <i>International Journal of Child & Family Welfare</i> , 9 (3), 131-144.	2	1	2	1	2	2	2	12	US	non-experimental process evaluation	n/a



	Reference	Research rationale	Research design	Sampling	Data collection	Data analysis	Interpretation and reporting of results	Credibility of conclusions	Overall score	Country of study	Study methods	Maryland score if appropriate
19	Crampton, D., & Jackson, W. L. (2007). Family group decision making and disproportionality in foster care: a case study. <i>Child Welfare Journal</i> , 86 (3), 51-69.	2	3	3	3	2	2	1	10	US	Impact evaluation non-comparable groups	2
21	Darlington, Y., Healy, K., Yellowlees, J., & Bosly, F. (2012). Parents' perceptions of their participation in mandated family group meetings. <i>Children and Youth Services Review</i> , 34 (2), 331-337.	2	0	0	1	1	2	1	7	Australia	Qualitative interviews	n/a
23	Jones, L.P., & Finnegan, D. (2004). Family unity meetings: decision making and placement outcomes. <i>Journal of Family Social Work</i> , 7 (4), 23-43.	1	1	2	1	1	1	2	9	US	Retrospective descriptive case record analysis	n/a
28	Hayden, C. (2009). Family group conferences: are they an effective and viable way of working with attendance and behaviour problems in schools? <i>British Educational Research Journal</i> . 35 (2), 205-220.	2	2	1	1	2	2	2	12	UK	Quasi-experimental matched comparisons	3
31	Hipple, N. K., & Mcgarrell, E. F. (2008). Comparing Police-And Civilian-Run Family Group Conferences. <i>Policing-An International Journal Of Police Strategies & Management</i> , 31 (4), 553-577.	2	1	2	2	2	1	2	12	US	Quasi-experimental	3
32	Hipple, N. K., Gruenewald, J. & McGarrell, E. F. (2014). Restorativeness, Procedural Justice, and Defiance as Predictors of Reoffending of Participants in Family Group Conferences. <i>Crime & Delinquency</i> , 60 (8), 1131-1157.	2	1	2	1	2	1	2	11	US	Non-experimental longitudinal	2
33	Holland, S., & O'Neill, S. (2006). 'We had to be there to make sure it was what we wanted' - Enabling children's participation in family decision-making through the family group conference. <i>Childhood-A Global Journal Of Child Research</i> , 13 (1), 91-111.	1	0	1	0	0	1	0	3	Wales	Exploratory qualitative study	n/a



	Reference	Research rationale	Research design	Sampling	Data collection	Data analysis	Interpretation and reporting of results	Credibility of conclusions	Overall score	Country of study	Study methods	Maryland score if appropriate
35	Jeong, S., McGarrell, E. F., Hipple, N. K. (2012). Long-term impact of family group conferences on re-offending: the Indianapolis restorative justice experiment. <i>Journal of Experimental Criminology</i> , 8 (4), 369-385.	2	2	2	1	1	2	2	12	US	Randomised control trial long-term follow up	5
36	Johansen, S. (2014). Psycho-social processes and outcomes of family group conferences for long-term social assistance recipients. <i>British Journal of Social Work</i> , 44 (1), 145-162.	2	1	0	0	1	2	1	7	Norway	Qualitative interview study	n/a
41	Malmberg-Heimonen, I. & Johansen, S. (2014). Understanding the longer-term effects of family group conferences. <i>European Journal of Social Work</i> , 17 (4), 556-571.	2	2	2	1	1	2	2	12	Norway	RCT with 12 month follow-up	5
44	Maxwell, G., & Kingi, V. (2001). Differences in How Girls and Boys Respond to Family Group Conference: Preliminary Research Results. <i>Social Policy Journal of New Zealand/Te Puna Whakaaro</i> , 17	1	1	1	1	2	2	1	9	New Zealand	Large scale retrospective study	n/a
45	Maxwell, G., Robertson, J., & Kingi, V. (2002). Achieving the Diversion and Decarceration of Young Offenders in New Zealand. <i>Social Policy Journal of New Zealand/Te Puna Whakaaro</i> , 19	1	1	2	1	0	1	1	7	New Zealand	Retrospective file case review	n/a
48	McGarrell, E. F., & Hipple, N. K. (2007). Family group conferencing and re-offending among first-time juvenile offenders: The Indianapolis experiment. <i>Justice Quarterly</i> , 24 (2), 221-246.	2	2	2	1	1	2	2	12	US	RCT	5
49	Morris, A. (2000). Family group conferences in New Zealand: reconstructions findings from a six year follow up study. <i>Prison Service Journal</i> . 128 , 13-16.	1	1	0	1	1	1	1	6	New Zealand	Retrospective follow up 6 years post FGC	n/a



	Reference	Research rationale	Research design	Sampling	Data collection	Data analysis	Interpretation and reporting of results	Credibility of conclusions	Overall score	Country of study	Study methods	Maryland score if appropriate
50	Mutter R., Shemmings, D., Dugmore, P., & Hyare, M. (2008). Family group conferences in youth justice. <i>Health and Social Care in the Community</i> , 16 (3), 262-270.	1	1	0	1	1	1	0	5	UK	Non-experimental pre-post test design	2
53	O'Shaughnessy R., Collins, C., & Fatimilehin I. (2010). Building bridges in Liverpool: exploring the use of family group conferences for black and minority ethnic children and their families. <i>British Journal of Social Work</i> , 40 (7), 2034-2049.	1	1	1	0	0	0	1	4	UK	Non-experimental design	2
54	Pennell J., Edwards M., & Burford G. (2010). Expedited family group engagement and child permanency. <i>Children and Youth Services Review</i> , 32 (7), 1012-1019.	2	1	2	2	1	2	2	12	US	Quasi-experimental	3
55	Pennell, J. (2006). Restorative practices and child welfare: toward an inclusive civil society. <i>Journal of Social Issues</i> , 62 (2), 259-279.	1	1	1	2	1	2	2	10	US	Non-experimental	n/a
58	Pennell, J., & Burford, G. (2000). Family group decision making: protecting children and women. <i>Child Welfare</i> , 79 (2), 131-158.	2	1	1	1	0	1	2	8	Canada	Quasi-experimental	3
60	Pountney, K. (2005). <i>Education and early intervention projects: an evaluation of the first year</i> . Essex County Council. Family Group Conference Service	1	0	0	1	0	1	1	4	UK	Non-experimental	2
61	Pugh, R. (2002). A family group conference pilot project: evaluation and discussion. <i>Practice: Social Work in Action</i> , 14 (2), 45-58.	1	0	0	1	0	1	1	4	UK	Non-experimental	2
65	Sheets, J., Wittenstrom, K., Fong, R., James, J., Tecci, M., Baumann, D.J., & Rodriguez, C. (2009). Evidence-based practice in family group decision-making for Anglo, African American and Hispanic families. <i>Children and Youth Services Review</i> , 31 (11), 1187-1191.	2	1	1	1	2	2	2	11	US	Quasi-experimental	4



	Reference	Research rationale	Research design	Sampling	Data collection	Data analysis	Interpretation and reporting of results	Credibility of conclusions	Overall score	Country of study	Study methods	Maryland score if appropriate
68	Sundell, K., & Vinnerljung, B. (2004). Outcomes of Family Group Conferencing in Sweden. A 3-Year Follow-Up. <i>Child Abuse & Neglect</i> , 28(3), 267-287.	2	2	2	1	2	2	2	13	Sweden	Quasi-experimental	4
70	Walton, E., Roby, J., Frandsen, A., & Davidson, R. (2003). Strengthening At-Risk Families by Involving the Extended Family. <i>Journal of Family Social Work</i> , 7(4), 1-21.	1	1	1	1	1	1	1	7	US	Process & impact evaluation	n/a
71	Wang, E.W., Wang, Lambert, M. C., Johnson, L. E., Boudreau, B., Breidenbach, R, & Baumann, D. (2012). Expediting permanent placement from foster care systems: the role of family group decision-making. <i>Children and Youth Services Review</i> , 34(4), 845-850.	1	2	2	1	2	2	2	12	US	Quasi-experimental	4
72	Weigensberg, E. C., Barth R., Guo S. (2009). Family group decision making: a propensity score analysis to evaluate child and family services at baseline and after 36-months. <i>Children and Youth Services Review</i> , 31(3), 383-390.	2	2	2	2	2	1	2	13	US	Quasi-experimental	4
73	Zernova M.(2007). Aspirations of restorative justice proponents and experiences of participants in family group conferences. <i>British Journal of Criminology</i> , 47(3), 491-509.	1	0	0	0	1	1	1	4	UK	Non-experimental	n/a



*The criteria for each level of the Maryland Scientific Methods Scale (SMS) are:

1. Correlation between a prevention programme and a measure of crime at one point in time (e.g. areas with CCTV have lower crime rates than areas without CCTV);
2. Measures of crime before and after the programme, with no comparable control conditions (e.g. crime decreased after CCTV was installed);
3. Measures of crime before and after the programme in experimental and control conditions (e.g. crime decreased after CCTV was installed in an experimental area, but there was no decrease in crime in a comparable area);
4. Measures of crime before and after in multiple experimental and control units, controlling for the variables that influence crime (e.g. victimisation of premises under CCTV surveillance decreased compared with victimisation of control premises, after controlling for features of premises that influenced their victimisation);
5. Random assignment of programme and control conditions to units (e.g. victimisation of premises randomly assigned to have CCTV surveillance decreased compared with victimisation of control premises).

The authors of the SMS suggest that confidence in intervention results is highest at level 5 and level 3 should be the minimum level required to achieve reasonably accurate results.



Appendix D – quality scores for reviews

Generic Questions	Quality Appraisal	Ratings 0-2
REVIEW METHOD	Comprehensive review of previous research and justification for reviewing multiple sources of data rather than conducting new primary research (including reference to other reviews/metas)	
	Clear identification of the research question and study aims, its context and objectives.	
	Was the review systematic? i.e. was there a clear process that is supported by other evidence	
	Were appraisal tools/ methods piloted, including search?	
	Reliability - triangulation of search, coding and analysis/appraisal - were multiple researchers used and agreement rates provided? How were differences in coding/scores resolved?	
	Subscore:	
SEARCH STRATEGY	Detailed explanation of search strategy and boundaries, including explanation of why key terms & synonyms were used (i.e. could the search be easily replicated to find similar results / update)	
	Sources - were a wide range of databases and websites searched covering multiple sources of data?	
	If subsequent searches were performed on references within the initial search or contact with experts, are there details of the process and criteria used to propose inclusion?	
	External validity (robustness of search)- are the databases used likely to ensure a comprehensive search with maximised inclusion and limited bias? If there are few negative findings (for effect studies) have unpublished articles been sourced?	
	Non-English language studies - if not included is there a detailed explanation (e.g. phenomenon specific to UK or cross-cultural studies would confound results)?	
	Accounts for or acknowledges publishing bias towards significant results	
	Was the search timeline explicitly stated and appropriate to the scope of the research question, considering the number of relevant studies published?	



Generic Questions	Quality Appraisal	Ratings 0-2
	Subscore:	
DATA COLLECTION (SIFT)	Description of studies and how and why chosen - details of pre-determined sift criteria that could be replicated	
	Description of population(s) of interest and how sample selection (s) relates to it and allows comparisons to be made	
	If there are too many studies to reasonably include in a review or meta, was a random sample chosen through an explicit system?	
	Description of methods to maximise inclusion/ secure representative coverage and limit potential for sample bias	
	Did the search criteria give sufficient attention to ethical issues - to the extent that it limits potential for bias and the possibility of skewing the type of studies included?	
	Subscore:	
QUALITY APPRAISAL	Validity of results - are opposing viewpoints included and discussed; are conclusions plausibly based on the data and not researcher's pre-conceptions (e.g. has the researcher critically reflected on own biases and influence and research skills?)	
	Explicit analytic procedure for processing raw data into results / themes that could be repeated with a similar methodology. Were the methods employed (eg, statistical tests/ models for quantitative research) appropriate?	
	Reliability - was there triangulation of data analysis (e.g. multiple scorers or coders)	
	Quality appraisal tool - robust with detailed explanation (or copy as appendix)	
	Marking criteria included considerations of ethics, researcher bias, comparability of any control groups, context and reliability of data collection (included representativeness of sample), quality of analyses, validity of results, and credibility of conclusions	
	Open explanation of rules/tool for classification of variables (e.g. different types of treatments / interventions)	
	Have the authors identified all important confounding factors and adequately taken them into account in the analysis? Eg, for quantitative research: restriction in design and techniques.	
	Subscore:	
DATA ANALYSIS / SYNTHESIS - QUANTITATIVE	Explicit analytic procedure for processing raw data into results / themes that could be repeated with a similar methodology.	
	Have the authors identified all important confounding factors and adequately taken them into account in the analysis? Eg, for quantitative research: restriction in design and techniques, eg, modelling, stratified-, regression-, or sensitivity analysis to	
	Coding of variables - openly explains procedure and specifies categories and units for scales	
	Codes quality of studies (and research designs)	



Generic Questions	Quality Appraisal	Ratings 0-2
	Has multiple regression analysis been performed on independent/moderator variables to separate out effects (when many variables)?	
	Were the methods employed (eg, statistical tests/ models) appropriate? E.g. using 'd' for effect sizes of categorical variables and 'r' for continuous variables	
	Has sample size been taken into account, either by weighing studies based on sample size or giving equal sizes to all studies?	
	Were details given of calculation of effect sizes (e.g. from means and standard deviations presented in the studies)?	
	Describes procedure for examining the distribution of effect sizes and analysing the impact of moderating variables, inc. details of statistical tests	
	Subscore:	
QUALITATIVE SYNTHESIS	Meta-ethnography - detailed description of qualitative analyses	
	Discussion of how error or bias may have arisen in design/ data collection/ analysis and how addressed, if at all	
	Have the authors identified all important confounding factors and adequately taken them into account in the analysis?	
	Search was exhaustive and analysis reached 'data saturation' (i.e. looking at new studies won't add to the knowledge base)	
	Common themes are grouped together but individual nuances preserved	
	Subscore:	
INTERPRETATION AND REPORTING OF RESULTS	Are the main results presented clearly and with reference to confidence intervals etc if appropriate?	
	Findings/ conclusions 'make sense' (have a coherent logic) and clear discussion of how they were derived and evidence to support them	
	Discussion of the mechanism through which a causal relationship might occur	
	Identification of patterns of association/ linkages, with descriptions of divergent positions/ multiple perspectives and any anomalous/negative cases	
	Discussion of how error or bias may have arisen in design/ data collection/ analysis and how addressed, if at all - limitations that may affect generalisability	
	Were effect sizes presented clearly as histograms, forest plots etc., if appropriate?	
	Discussion of implications of findings for policy or practice; identification of new avenues of research (e.g. potential new moderators)	
	Discussion of how context may shape an intervention's effects (e.g. does it work on some groups and not on others; are significant effects found)	
	Subscore:	



Generic Questions	Quality Appraisal	Ratings 0-2
CREDIBILITY OF CONCLUSIONS	Discussion/ evidence of the main assumptions/ hypotheses/ theoretical ideas on which the research was based and how these affected the form, coverage or output of the research	
	Conclusions presented are supported by study findings and previous research and theory (where appropriate).	
	Evidence of openness to new/ alternative ways of viewing subject/ theories/ assumptions	
	Subscore:	



Quality assessment summary for reviews

All criteria marked on a scale of 0 – 2

	Reference	Are review methods robust? (2)	Are search strategies sufficiently comprehensive? (3)	Are sift criteria reported in sufficient detail? (4)	Has quality appraisal of studies been adequately reported? (5)	Have the results of quantitative studies been adequately synthesised? (6)	Is the synthesis of qualitative studies adequate? (7)	Is the interpretation and reporting of results presented clearly? (8)	Are the study conclusions credible? (9)	Overall score	Country of origin	Review method
5	Barnsdale L. & Walker M. (2007). <i>Examining the use and impact of family group conferencing</i> . Report for the Scottish Executive.	1	1	1	1	0	2	1	1	8	Scotland	Literature review
10	Bradshaw, W., & Roseborough, D, (2005). An empirical review of family group conferencing in juvenile offenses. <i>Juvenile And Family Court Journal</i> , 56(4), 21-28.	1	1	1	1	1	1	1	2	9	US	Meta analysis
13	Chand A., & Thoburn, J. (2005). Child and family support services with minority ethnic families: what can we learn from research? <i>Child and Family Social Work</i> , 10(2), 169-178.	0	0	0	0	0	1	1	1	3	UK	Literature review
18	Crampton, D. (2007). Research review: family group decision-making: a promising practice in need of more programme theory and research <i>Child and Family Social Work</i> , 12(2), 202-209.	1	1	0	0	0	0	1	1	4	US	Literature review
20	Crea T. M., & Berzin, S. C. (2010). Family involvement in child welfare decision-making: strategies and research on inclusive practices. <i>Journal of Public Child Welfare</i> , 3(3), 305-327.	1	0	0	0	0	0	1	1	3	US	Literature review
26	Frost, N., Abram, F., & Burgess, H. (2014). Family group conferences: evidence, outcomes and future research. <i>Child & Family Social Work</i> , 19(4), 501-507.	1	0	0	1	0	1	1	1	5	UK	Literature review
40	Levine, M. (2000). The family group conference in the New Zealand Children, Young Persons, and Their Families Act of 1989: review and evaluation. <i>Behavioral Sciences and the Law</i> , 18(4), 517-556.	1	0	0	0	0	0	1	2	4	New Zealand	Literature review



	Reference	Are review methods robust? (2)	Are search strategies sufficiently comprehensive? (3)	Are sift criteria reported in sufficient detail? (4)	Has quality appraisal of studies been adequately reported? (5)	Have the results of quantitative studies been adequately synthesised? (6)	Is the synthesis of qualitative studies adequate? (7)	Is the interpretation and reporting of results presented clearly? (8)	Are the study conclusions credible? (9)	Overall score	Country of origin	Review method
+1	Sherman, L.W., Strang, H., Mayo-Wilson, E., Woods, D.J.(2015). Are Restorative Justice Conferences Effective in Reducing Repeat Offending? Findings from a Campbell Systematic Review. <i>Journal of Quantitative Criminology</i> , 31 , 1-24.	2	2	1	2	2	2	2	2	15	UK	Systematic Review

