

Technical annex: Guidance recommendation standards

MENTORING AND BEFRIENDING PRACTICE GUIDE: FOR CARE EXPERIENCED CHILDREN AND YOUNG PEOPLE AND THOSE AT RISK OF ENTERING CARE

Introduction

This document sets out our approach for translating the findings of the systematic review and meta-analysis into actionable recommendations and key principles. The purpose of using these guidance recommendation standards is to help individuals involved in the design, commissioning, delivery and evaluation of children's social care services to understand the quality, reliability, and validity of the evidence recommended in the Practice Guides.

In line with best practice, these guidance recommendation standards consider existing validated tools for assessing the quality of evidence. Specifically, these guidance recommendation standards incorporate criteria items from the EMMIE Framework¹ and that of AMSTAR 2² (A Measurement Tool to Assess Systematic Reviews).

We have chosen to use relevant criteria items from the EMMIE and AMSTAR 2 frameworks because in addition to assessing whether an intervention works (effectiveness), EMMIE also considers how the intervention can be implemented and factors that can influence its implementation. The EMMIE framework is applicable to the kind of questions that our systematic reviews and Practice Guides seek to answer, which are 'what works and for whom', how interventions, programmes or services work, and the contextual factors that influence how

¹ Johnson, S. D., Tilley, N., & Bowers, K. J. (2015). Introducing EMMIE: An evidence rating scale to encourage mixed-method crime prevention synthesis reviews. *Journal of experimental criminology*, 11, 459-473.

² Shea, B.J., Reeves, B.C., Wells, G., Thuku, M., Hamel, C., Moran, J., Moher, D., Tugwell, P., Welch, V., Kristjansson, E., & Henry, D.A. (2017). AMSTAR 2: a critical appraisal tool for systematic reviews that include randomized or non-randomised studies of healthcare interventions or both. *The British Medical Journal: Research Methods and Reporting*, 358. <https://www.bmj.com/content/358/bmj.j4008>.



interventions work, as well as evidence of how to implement them in practice. The AMSTAR 2 critical appraisal tool is a widely used framework that allows us to assess the confidence in the systematic review methodology.

About the EMMIE framework

The EMMIE framework was developed by the Jill Dando Institute at the University College of London, and it was primarily developed for use by the What Works Centre for Crime Reduction to identify and rate the certainty of evidence from systematic reviews. It was also used by the former What Works for Children's Social Care to capture the nuance and complexity of evidence in the Children's Social Care sector. EMMIE is an acronym for:

- **Effect:** the difference an intervention makes based on existing evidence of its evaluation. It is an estimate of the overall effect size of an intervention and can be used to determine the confidence that should be placed on that estimate.
- **Mechanisms/mediators:** this refers to how the intervention, programme, or service produces effect. In other words, what makes the intervention work.
- **Moderator:** the factors or things (e.g. the context, setting, or time of intervention delivery) that might affect whether an intervention works.
- **Implementation:** refers to evidence about how to implement the intervention, programme, or service.
- **Economic impact:** refers to the cost-effectiveness of the programme, intervention, or service. In other words, the cost of implementing the intervention or using the service.

The EMMIE framework uses a 5-point rating scale to assess the quality of evidence from systematic reviews. Further information about the EMMIE framework is available [here](#).³

About the AMSTAR 2 critical appraisal tool

The AMSTAR 2 tool is a critical appraisal framework for assessing the methodological quality of systematic reviews that include randomised and/or non-randomised studies of interventions. It covers 16 aspects of a systematic review, full details of which can be found [here](#).⁴ In addition to appraising aspects of a systematic review, such as whether the review protocol was registered before commencing the review, appropriateness of the search strategy and search terms, and the selection criteria, the AMSTAR 2 tool gives prominence to risk of bias assessment and how it can influence the validity of a systematic review finding and its conclusions. The tool uses up to 3-point scales to assess the methodological quality of systematic reviews.

³ See: <https://discovery.ucl.ac.uk/id/eprint/10141017/1/5.Sidebottom%26Tilley.pdf>

⁴ See: <https://amstar.ca/docs/AMSTAR-2.pdf>



Our evidence rating

Evidence rating refers to a systematic approach for grading and assessing the quality of evidence that is produced from a research study, a systematic review, or expert opinion. Evidence rating enables us to grade the strength of evidence for key principles and recommendations in the Practice Guides, so that users of the Guide will have a better understanding of the strength and reliability of the recommendations, and the significant impact they can have on early intervention and social work practice.

Our evidence rating consists of criteria items from four domains of the EMMIE framework, which are Effect, Mechanisms/Mediators, Moderators, and Implementation, as well as all 16 items of the AMSTAR 2 critical appraisal tool. To assess the strength of evidence, our evidence rating uses a 3-point scale, which are 'Yes', 'Unsure/Can't tell' and 'No'. Each domain is aligned to a particular research question of the systematic review that is being appraised.

In order to rate the certainty of evidence for each research question, each item of the EMMIE framework and AMSTAR 2 appraisal tool was assigned a score of '1' (Yes) if the item has been satisfactorily addressed in the systematic review report. We then provided a total rating of the certainty of evidence, which is the sum of the individual items of each domain. This is illustrated in the table below.

Domain	Aligned Practice Guide research question	Total score
Review methodology (AMSTAR 2)	Assesses the quality of the systematic review methodology.	16
Assessment of effectiveness	RQ1: Which interventions work to improve outcomes?	9
Mechanism/Mediators	RQ3: How and why does the intervention work?	6
Moderators	RQ2: For whom does the intervention work? RQ3: How and why does the intervention work?	6
Implementation	RQ4: What are the enablers and barriers of implementing the intervention? RQ5: What are the users'/practitioners' views and experiences of the intervention?	4

Then, we provided an overall score of the evidence rating, which is a combination of scores for *Review methodology* and the individual domains relevant to each research question/finding, as specified in the table above. For example, when assessing certainty of evidence for research question (RQ) 1, we combine the scores for the domains 'review methodology' and 'assessment of effectiveness'. The maximum score for each research question is given below:



Domain and aligned Practice Guide research question	Overall score
Review methodology	16
Assessment of effectiveness	9 + 16 = 25
Mechanism/Mediators	6 + 16 = 22
Moderators	6 + 16 = 22
Implementation	4 + 16 = 20

Assessing effectiveness of interventions

Meta-analysis

Where studies have been combined in a meta-analysis, we apply all the items of the ‘Effectiveness’ domain of the EMMIE framework to assess effectiveness of interventions. We assign a score of ‘1’ (Yes) if two or more randomised controlled trials and/or quasi-experimental studies were included in the meta-analysis. Also, we assign a score of ‘1’ (Yes) if the review authors have conducted an analysis of heterogeneity in effect sizes using appropriate methods, such as calculating a statistical test for heterogeneity (the I^2 value), conducting sub-group analysis or meta-regression, and visual evaluations of forest plots. Where an analysis of heterogeneity in effect sizes is conducted, we assign a score of ‘1’ (Yes) if the I^2 value is less than or equal to 60%⁵. Other items of the ‘Effectiveness’ domain of the EMMIE framework can be found [here](#).⁶

In addition, we consider the appropriateness of the design used for conducting included studies. In our systematic reviews, we only include randomised controlled trials and quasi-experimental impact evaluations to assess intervention effectiveness. We also consider the risk of bias assessment score for each of the studies included in meta-analysis. The recommendations in the Practice Guides are only based on studies which have scored at least ‘moderate’ or ‘some concern’ on risk of bias assessment. We do not make recommendations from studies that have scored ‘high’ on risk of bias assessment.

Furthermore, we use the Foundations’ Guidebook criteria on sample size to determine whether a sufficient sample was used to assess intervention effectiveness:

- For included randomised controlled trials and quasi-experimental studies, the analysis sample (i.e. the sample upon which the main analysis of outcomes or programme

⁵ Higgins, J.P.T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M.J., Welch, V.A. eds (2023). Cochrane Handbook for Systematic Reviews of Interventions version 6.4 (updated August 2023).

⁶ See: <https://pdfs.semanticscholar.org/aa35/a06165b3daa1ededcd6154d0627fd1028cac.pdf>



effectiveness is conducted) must contain at least 20 participants in each group (i.e., the intervention and control group) to receive a rating of ‘strong’

- Where a study contains at least 20 participants in the intervention group, but less than 20 in the comparison group, we assign a rating of ‘good’
- A rating of ‘promising’ is assigned if a study contributing to a narrative synthesis finding contains less than 20 in the treatment group, but 20 or over in the control group.

Narrative synthesis

Where there is no meta-analysis, but a narrative synthesis of findings is present, we have adapted the validity assessment approach recommended by EPPI⁷ to rate the certainty of the evidence. This includes the following:

- The methodological soundness (trustworthiness): this is judged based on the risk of bias score of each study included in narrative synthesis for the assessed outcome. As stated above, recommendations in the Practice Guides are only based on studies which have scored at least ‘moderate’ or ‘some concern’ on risk of bias. Also, as part of this criteria, we determine whether a sufficient sample was used to assess intervention effectiveness, using Foundations’ Guidebook criteria on sample size described above.
- The appropriateness of the study design in answering the review question (appropriateness): this assesses whether the design used for conducting the study was appropriate to answer the research question(s). As stated above, in our systematic reviews, we only include randomised controlled trials and quasi-experimental impact evaluations to assess intervention effectiveness.
- The study relevance (relevance): this assesses relevance of the context where the included studies were conducted. For the purposes of transferability, studies must be conducted in high-income countries with comparable child welfare systems to the UK. These include, but are not limited to, all the four nations of the UK (England, Scotland, Wales, Northern Ireland), the United States, Australia, Canada, the Republic of Ireland and other European countries.
 - This criterion also assesses the relevance of the criteria used for selecting studies to be included in the systematic review. The systematic review authors must provide adequate information about the inclusion and exclusion criteria and provide more details about the intervention(s) that were delivered, as well as the population and outcomes of interest.
- An assessment of the overall weight of the evidence which the study provides. The first three criteria above contribute to the assessment of the study weight. The overall weight of each study contributing to a finding will determine the certainty of evidence rating for that finding.

⁷ Popay, J., H. Roberts, A. Sowden, M. Petticrew, L. Arai, N. Britten, M. Rodgers, K. Roen and S. Duffy (2006) *Guidance on the Conduct of Narrative Synthesis in Systematic Reviews: Final Report*. Swindon: ESRC Methods Programme.



Strength of evidence rating: Quantitative evidence

To assess the strength of the evidence relating to the effectiveness of interventions, we use the approach presented in the table below:

Figure 1. Approach for rating the strength of quantitative evidence ([go to accessibility text](#))

STRONG EVIDENCE	A rating of ' strong ' is given if the evidence is from a meta-analysis or narrative synthesis of at least two randomised controlled trials and/or quasi-experimental studies that were conducted in the UK or a comparable high-income country and have scored low on risk of bias assessment, with a minimum sample size of 20 in each group (i.e., the intervention and comparison group), and demonstrates effectiveness of the intervention(s).
GOOD EVIDENCE	A rating of ' good ' is given if the evidence is from a meta-analysis or narrative synthesis of at least two randomised controlled trials and/or quasi-experimental studies that were conducted in the UK or a comparable high-income country and have scored at least moderate on risk of bias assessment, with at least 20 participants in the intervention group and less or more than 20 participants in the comparison group, and demonstrates efficacy of the intervention(s).
PROMISING EVIDENCE	A rating of ' promising ' is given where the evidence is from one randomised controlled trial or quasi-experimental study that was conducted in the UK or a comparable high-income country and has scored low or moderate on risk of bias assessment, with less or more than 20 participants in each group (i.e., the intervention and comparison group) and demonstrates efficacy of the intervention(s).

Strength of evidence rating: Qualitative evidence

Within the systematic reviews and Practice Guides, we aim to assess how and why interventions work to improve outcomes, for whom, the enablers and barriers to successful implementation of interventions, and user or beneficiary perspectives and experiences of the intervention. Thus, qualitative evidence may be included to address these features of the systematic review and Practice Guide. Where there is qualitative evidence, we use the following approach to determine certainty of the evidence:

- 1) The methodological quality of the individual studies as reported in the review report. We expect that review authors will use appropriate validated critical appraisal tools to assess methodological quality of included qualitative studies.
- 2) The coherence of each finding.

To assess coherence of each qualitative finding, firstly, at least two reviewers independently extract a set of statements that cover key findings of interest from thematic analysis of qualitative findings. This is done for each qualitative research question (for example, enablers and barriers of



implementing the intervention and users’/practitioners’ views and experiences of intervention delivery). Then, for each finding, reviewers make a note of the number of studies that contribute to that finding. Reviewers then discuss among themselves and agree key findings of interest and the number of studies that contributed to that finding, as well as the relevance of the context in which the study was conducted. Reviewers then assign an overall evidence rating to the specific finding (i.e. promising, good, or strong evidence of certainty), with a statement explaining the rating.

In rating the strength of qualitative evidence, the following approach is used:

Figure 2. Approach for rating the strength of qualitative evidence ([go to accessibility text](#))

STRONG EVIDENCE	A rating of ‘ strong ’ is given if the evidence is from at least five studies that were conducted in the UK context and have scored high on methodological quality.
GOOD EVIDENCE	A rating of ‘ good ’ is given if the evidence is from at least two studies that were conducted in the UK context and have scored at least moderate on methodological quality.
PROMISING EVIDENCE	A rating of ‘ promising ’ is given if the evidence is from only one study conducted in the UK context, and have scored high or moderate on methodological quality.

Go to the Mentoring & Befriending Practice Guide:

<https://foundations.org.uk/toolkit/practice-guides/mentoring-and-befriending>

Find out more about the series of Practice Guides: foundations.org.uk/practice-guides



Accessibility text

Figure 1. Quantitative evidence ratings

The image shows three levels of evidence ratings, each with a label and an accompanying description. The information is presented vertically:

1. **Strong evidence** (darkest blue oval label):
 - A rating of 'strong' is given if the evidence is from a meta-analysis or narrative synthesis of at least two randomised controlled trials and/or quasi-experimental studies that were conducted in the UK or a comparable high-income country and have scored low on risk of bias assessment, with a minimum sample size of 20 in each group (i.e., the intervention and comparison group), and demonstrates effectiveness of the intervention(s).
2. **Good evidence** (light blue oval label):
 - A rating of 'good' is given if the evidence is from a meta-analysis or narrative synthesis of at least two randomised controlled trials and/or quasi-experimental studies that were conducted in the UK or a comparable high-income country and have scored at least moderate on risk of bias assessment, with at least 20 participants in the intervention group and less or more than 20 participants in the comparison group, and demonstrates efficacy of the intervention(s).
3. **Promising evidence** (pale blue oval label):
 - A rating of 'promising' is given where the evidence is from one randomized controlled trial or quasi-experimental study that was conducted in the UK or a comparable high-income country and has scored low or moderate on risk of bias assessment, with less or more than 20 participants in each group (i.e., the intervention and comparison group) and demonstrates efficacy of the intervention(s).

[Click here to return to report](#)

Figure 2. Qualitative evidence ratings

The image consists of three levels of evidence ratings, each represented by a different label and description. The ratings are displayed in a vertical format:

1. **Strong evidence** (darkest blue oval label):
 - A rating of 'strong' is given if the evidence is from at least five studies that were conducted in the UK context and have scored high on methodological quality.
2. **Good evidence** (light blue oval label):



- A rating of 'good' is given if the evidence is from at least two studies that were conducted in the UK context and have scored at least moderate on methodological quality.

3. **Promising evidence** (pale blue oval label):

- A rating of 'promising' is given if the evidence is from only one study conducted in the UK context and have scored high or moderate on methodological quality.

[Click here to return to report](#)