

Appendix 4 - Supplementary table of study evaluations including quality assessments and relative contributions to themes

Article	Theme	1					2		3			4			5			6		Number of sub-themes ^a	QualSyst score ^b	GRADE score ^c	
	Subtheme	1.1	1.2	1.3	1.4	1.5	2.1	2.2	3.1	3.2	3.3	4.1	4.2	4.3	5.1	5.2	5.3	6.1	6.2				
Aman et al. [58]							•	•		•										3	0.80	Mod	
Sowney and Barr [63]		•	•		•					•									•	5	0.60	High	
Carlson et al. [73]		•	•		•		•					•								4	0.72	Low	
Ferguson and Murphy [59]							•								•					2	1.00	Mod	
Ferguson et al. [64]					•		•	•	•	•						•				7	0.90	Mod	
Fisher et al. [65]		•			•	•					•									3	0.60	Low	
Fish et al. [60]			•		•		•	•		•				•					•	•	8	0.65	High
Goldsmith, et al. [67]		•	•				•									•	•	•		5	0.90	High	
Graham et al. [61]			•														•			2	0.90	Mod	
Hall et al. [68]							•													1	0.80	High	
Hart [65]		•			•	•											•			3	0.55	Mod	
Hoglund and Larsson [72]		•	•			•	•							•					•	•	7	0.75	High
Huneke et al. [62]		•			•		•	•											•		5	0.55	Mod
McCarthy 2010 [56]		•	•		•		•	•	•			•	•	•	•	•				10	0.80	High	
Rogers et al. [66]		•	•	•	•	•	•				•	•	•	•	•	•	•		•	10	0.95	High	
O'Dell et al. [70]		•		•											•				•	4	0.94	Very low	
Rose et al. [74]							•	•												2	0.70	Mod	
Walmsley et al. [75]		•			•	•	•	•		•			•							6	0.70	High	
Wiseman and Ferrie [76]			•	•	•	•		•		•	•			•						7	0.80	High	
Wood and Douglas [77]			•				•						•		•					4	0.88	Low	
Arcscott et al. [71]							•								•	•				3	0.91	Very low	
Ledger et al. [78]		•		•	•	•				•										4	0.94	Mod	
McCarthy 2009 [55]		•	•		•	•	•	•	•	•						•	•		•	10	0.80	High	
<i>Number of articles that mention each subtheme^d</i>		13	11	4	13	8	16	9	4	7	6	5	2	5	5	5	5	3	7				
<i>Frequency rating^e</i>		A	A	C	A	B	A	B	C	B	C	C	D	C	C	C	C	D	B				

^aNumber of different subthemes identified in each included study article as a measure of study sensitivity.

^bQualSyst scores were calculated as an assessment of quality for each study.[41]

^cThe GRADE framework was applied to assess the quality of the body of evidence for each study with categories: very low, low, mod (moderate) and high.[50]

^dNumber of included articles that describe each subtheme as a measure of effect size.

^eFrequency rating was determined by the number of articles that mentioned each subtheme: 'A' if they were identified by 10 or more articles, 'B' if identified by 7-9 articles, 'C' if identified by 4-6 articles, and 'D' if identified by 3 or less articles.