**SDC 2. National Institute of Health (NIH) quality assessment tool.**

| STUDY NAME AND YEAR | GOOD | FAIR | POOR |
|----------------------|------|------|------|
| Blanchard et al, 1989 |     | 1    |      |
| Cohen-Kettenis et al, 1997 | | 1 | |
| De Cuyper et al, 2006 | 1 | | |
| Jiang et al, 2018 | 1 | | |
| Kuiper et al, 1998 | | | 1 |
| Lawrence, 2003 | | 1 | |
| Lobato et al, 2006 | | 1 | |
| Oslo-Kennedy et al, 2018 | | 1 | |
| Smith et al, 2001 | | 1 | |
| Van DeGrift et al, 2018 | | 1 | |
| Weipjes et al, 2018 | | 1 | |
| Judge et al, 2014 | | | 1 |
| Vujovic et al, 2009 | | | 1 |
| Weyers et al, 2009 | 1 | | |
| Nelson et al, 2009 | 1 | | |
| Poudrier et al, 2019 | | 1 | |
| Laden et al, 1998 | | 1 | |
| Bouman et al, 1988 | | 1 | |
| Johansson et al, 2010 | 1 | | |
| Kreger et al, 2001 | | | 1 |
| Papadopulos et al, 2017 | 1 | | |
| Pfafflin, 1993 | | 1 | |
| Rehman et al, 1999 | 1 | | |
| Song et al, 2010 | | 1 | |
| Zavlin et al, 2018 | 1 | | |
| Garcia et al, 2004 | | 1 | |
| Imbimbo et al, 2009 | | 1 | |
| **TOTAL** | 27 | 13 | 9 |