All three versions of SPARX are largely similar in terms of content, but each features a tailored script which makes the story more relevant as a targeted treatment approach for depression (SPARX), universal preventative approach for young people feeling down, angry or stressed (SPARX-R), or targeted for LGBT youth and the unique difficulties they face (Rainbow SPARX). It has recently been adapted and translated into Japanese (Yokomitsu et al., 2020) and is currently being redesigned for Nunavut youth (First Nations young people in Arctic Canada) (Bohr & Merry, 2016).

Format of Delivery and Main Features

SPARX is delivered in the format of seven sequential modules or levels, each of approximately 20-30 minutes duration, which can be accessed online or via a CDROM (Kuosmanen et al., 2017). The player controls an avatar across the seven distinct levels while interacting with characters who explain CBT concepts, teach the player skills and strategies, and give the player 'homework' to try these strategies while away from the game. Table 1 provides a summary of content within each module.

Table 1

Overview of SPARX modules, adapted from Lucassen et al., (2015)

Module	Main Content Covered
1 Cave Province: 'Finding Hope'	Introducing unhelpful thoughts The character and concept of Hope Controlled breathing Psycho-education about depression and the CBT model
2 Ice Province: 'Being Active'	Progressive muscle relaxation Communication skills Behavioural activation and activity scheduling

3 Volcano Province: 'Dealing with Emotions'	Listening skills Identifying strong emotions
4 Mountain Province: 'Overcoming Problems'	Introducing problem solving Recognising sparks (positive or helpful thoughts about you/your future)
5 Swamp Province: 'Recognising Unhelpful Thoughts'	Recognising various negative automatic thoughts
6 Bridgeland Province: 'Challenging Unhelpful Thoughts	Learning to challenge negative automatic thoughts
7 Canyon Province: 'Bringing it all together'	Recap of all skills Mindfulness Knowing when to ask for help

Mechanism of Change

SPARX is considered a 'serious game', in which education and behaviour change is the goal, alongside entertainment (Cheek et al., 2015). Serious

negative automatic thoughts that the situation elicited and the cognitive distortions which maintain a depressed mood. Therefore in CBT, individuals learn to address maladaptive behaviours and psychological distress by altering the cognitive processes and behaviours that sustain them (Beck, Rush, Shaw

The efficacy of SPARX as a clinical treatment for young people seeking help for low mood or depression has been demonstrated in a large randomised controlled non inferiority trial (RCT) (Merry et al., 2012), which reported pre- to post-intervention decreases in depressive symptoms for adolescents using SPARX, and equivalent outcomes between SPARX and face-to-face therapy (usual care).

Various qualitative studies of youth trialling or engaging with SPARX have reported high levels of satisfaction with the programme. Themes included

v6 Tm08871 0 595.32 841.92 reW* nBT/F3 12 Tf1 0 0 1 204.05 554.59 Tm0 g841.9(e)-3(m)3(

There is no current systematic synthesis of the effectiveness of SPARX in educational settings. This review therefore seeks to understand whether SPARX holds promise as an effective intervention for reducing depressive symptoms for youth when delivered through the educational setting context, such as during class time and with minimal supervision.

SPARX may have educational significance given the flexibility of its implementation and ability to extend the reach of standardised CBT to more young people, including those in the community whose access to support is limited by barriers inherent to depression itself, such as reduced motivation and help-seeking behaviour.

Digitalised SEMH interventions are particularly relevant for EP practice at the current time in which face-to-face practice is limited due to health and safety restrictions during COVID-19. The socio-emotional and mental health needs of young people continue to rise well beyond the capacity of current service provision, and questions remain as to how this can be addressed within education in the context of the pandemic and beyond (DfE, 2020; The Children's Society, 2020).

Review question

The primary aim of the current review is to address the question:

How effective is Ptex

ow ef f3 TJETQq0. TJET1(e)-3( )]TJETQ000(csvm)-3

				et al., 2010). 'Youth' is broadly defined as the 15-24 year age group (WHO, 2021)
2	Setting	SPARX must be delivered in an educational setting	SPARX not delivered in an educational setting	SPARX has not been reviewed in educational settings before
3	Type of Intervention	Study must have delivered a variant of SPARX intervention	Study did not deliver a variant of SPARX intervention	To allow the reviewer to critically evaluate the effectiveness of SPARX
4	Outcomes	One of the primary outcome measures are depressive symptoms	None of the primary outcome measure are depressive symptoms	To review the effectiveness of SPARX on depressive symptoms
5	Research design and methodology	Empirical data derived from quantitative studies with pre- and post-measures	Empirical data derived from qualitative studies or those missing quantitative data	Empirical data is sought for effect size calculation. An experimental design with pre- and post-measures is required to determine intervention impact
6	Type of publication	Studies published in an accessible peer-reviewed journal	Studies not published in an accessible peer-reviewed journal	Accessibility and peer-review is considered a minimum threshold of study quality and inclusion for this review

Figure 1

Study Selection Flow Chart

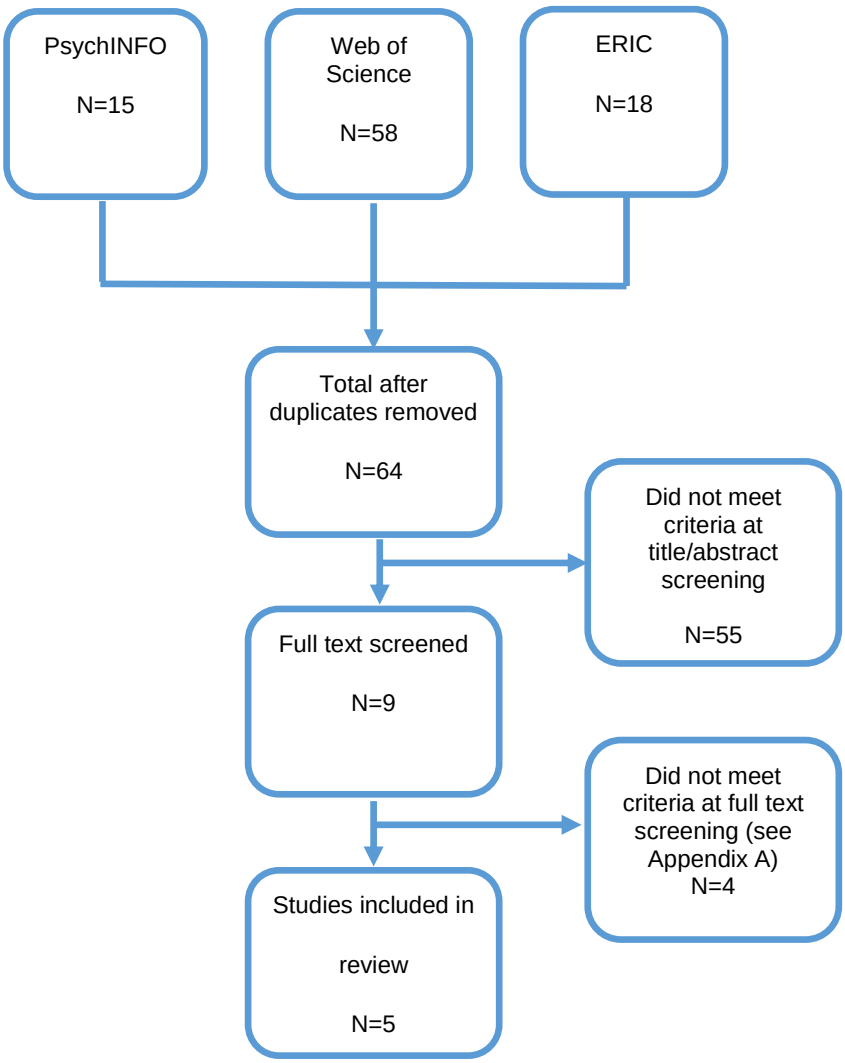

Table 5 summarises the WoE scores given to each study in this review. An adapted version of Kratochwill et al.'s (2003) protocol was used to assess WoE A for group designs. The sections omitted or modified and the accompanying rationale are detailed in Appendix B. Full details of the weight of evidence ratings and coding protocols are contained within the appendices.

Table 5

Overview of WoE ratings

Study	WoE A Methodological Quality	WoE B Methodological Relevance	WoE C Topic Relevance	WoE D Overall Weight of Evidence (Mean of A,B & C)
Fleming et al., 2012	Medium 2.25	High 2.5	Medium 2	Medium 2.25
Kuosmanen et al., 2017	Medium 1.75	High 2.5	Medium 2	Medium 2.08
Lucassen et al., 2015	Medium 1.67	Medium 1.5	High 2.5	Medium 1.89
Perry et al., 2017	High 3	High 2.5	High 2.5	High 2.67

Participants

A total of 841 young people from Ireland, The Netherlands, Australia and New Zealand and were included in the current review, ages ranging from 11 to 20 years old (see Table 6).

Participants were recruited from a variety of educational settings including mainstream secondary schools (selective and non-selective) and alternative education schools/programmes related to exclusion (for students at-risk of exclusion, currently excluded, transitioning out of exclusion and who have left school early). In addition to educational settings, one sample of participants (Lucassen et al., 2015) were also recruited by a sexual minority youth-led organization.

All studies reported the gender ratio of students in their sample, which was relatively equal, with one female-only study sample (Poppelaars et al., 2016). All participants in one study were reported as identifying as sexual minority youth (adolescents attracted to same sex, both sex, or who are questioning their sexuality) (Lucassen et al., 2015).

However, further reporting of demographic information was limited or inconsistent across all studies. Of the reported information, high cultural diversity was indicated in

Finally, there was no evidence of sufficient power in the non-randomised pre-post uncontrolled trial (Lucassen et al., 2015), increasing the likelihood of false-positive

Table 7

Table 7 presents a summary of the key findings. Effect sizes were calculated for impact of SPARX on depressive symptoms are expressed as standardised mean differences (Cohen's *d*). These were calculated from descriptive or univariate test stef*2 -3(re ca)-1 0 02cs,su-3(s)s

considerations of cultural specificity within educational settings when aiming to address the question of effectiveness.

Despite inconclusive findings of this review, SPARX is a pragmatic intervention with flexible delivery, particularly given the current context of online learning. SPARX therefore holds promise in bringing CBT-based intervention to a wider-reach of young people, including those in educational settings.

As a preliminary step towards dissemination, educational settings are recommended

References

Almeida, J., Johnson, R., Corliss, H., Molnar, B., & Azrael, D. (2009). Emotional Distress Among LGBT Youth: The Influence of Perceived Discrimination Based on Sexual Orientation. *Journal of Youth and Adolescence*, *38*, 1001–1014.

Angold, A., Costello, E.J., Messer, S.C., Pickles, A.,Winder, E., Silver, D. (1995). Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents: factor composition and structure across development. *Int. J. Methods Psychiatr. Res. 5*, 237–249.

Bech P, Rasmussen, N.A., Olsen, L.R., Noerho* nm

on SPARX, a Fantasy World Developed for New Zealand Youth With Depression. *Journal of Medical Internet Research*, *16*(2)

Cheek, C., Fleming, T., Lucassen, M. F. G., Bridgman, H., Stasiak, K., Shepherd, M., & Orpin, P. (2015). Integrating Health Behavior Theory and Design Elements in Serious Games. *JMIR Mental Health, 2*(2)

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.

Craighead, W. E., Smucker, M. R., Craighead, L. W., & Ilardi, S. S. (1998). Factor analysis of the Children's Depression Inventory in a community sample. *Psychological Assessment, 10*(2), 156-165.

Department for Education (DfE). (2020). *State of the nation 2020: children and young people's wellbeing.* Research Report. London DfE, The Stationery Office.

Ellis, L.A. (2004). Peers Helping Peers: The Effectiveness of a Peer Support Program in Enhancing Self-concept and Other Desirable Outcomes. University of Western Sydney, School of Psychology, Doctorate Thesis

Endicott, J., Nee, J., Ruoyong, Y., & Wohlberg, C. (2006). Pediatric quality of life enjoyment and satisfaction questionnaire (PQ-LES-Q): Reliability and validity. *Journal of the American Academy of Child and Adolescent Psychiatry, 45*(4), 401-407.

Fleming, T., Lucassen, M., Stasiak, K., Shepherd, M., & Merry, S. (2016). The impact and utility of computerised therapy for educationally alienated teenagers: The views of adolescents who participated in an alternative education-based trial. *Clinical Psychologist*, *20*(2), 94–102.

Horowitz, J., & Garber, J. (2006). The prevention of depressive symptoms in children and adolescents: A meta-analytic review. Journal of Consulting and Clinical Psychology, 74(3), 401-415. *Journal of Consulting and Clinical Psychology*, *74*, 401–415.

Kazdin, A. E., French, N. H., Unis, A. S., Esveldt-Dawson, K., & Sherick, R. B. (1983). Hopelessness, depression, and suicidal intent among psychiatrically disturbed inpatient children. *Journal of Consulting & Clinical Psychology, 51*(4), 504-510.

Kent, L., Vostanis, P., & Feehan, C. (1997). Detection of major and minor depression in children and adolescents: Evaluation of the mood and feelings questionnaire. *Journal of Child Psychology and Psychiatry, 38*(5), 565-573.

Kratochwill, T. R. (2003). Task Force on Evidence-Based Interventions in School Psychology. Retrieved February 14th, 2021, from http://www.indiana.edu/~ebi/documents/_workingfiles/EBImanual1.pdf

Kuosmanen, T., Fleming, T. M., Newell, J., & Barry, M. M. (2017). A pilot evaluation of the SPARX-R gaming intervention for preventing depression and improving wellbeing among adolescents in alternative education. *Internet Interventions*, *8*, 40–47.

Lucassen, M. F. G., Hatcher, S., Stasiak, K., Fleming, T., Shepherd, M., & Merry, S. N. (2013). The views of lesbian, gay and bisexual youth regarding computerised self-help for depression: An exploratory study. *Advances in Mental Health*, *12*(1), 22–33.

Lucassen, M. F. G., Merry, S. N., Hatcher, S., & Frampton, C. M. A. (2015). Rainbow SPARX: A Novel Approach to Addressing Depression in Sexual Minority Youth. *Cognitive and Behavioural Practic*e, 22(2), 203–216).

Merry, S. N., Stasiak, K., Shepherd, M., Frampton, C., Fleming, T., & Lucassen, M. F. G. (2012). The effectiveness of SPARX, a computerised self help intervention for adolescents seeking help for depression: Randomised controlled non-inferiority trial. *BMJ : British Medical Journal (Online)*, *344*.

Petticrew, M., & Roberts, H. (2003). Evidence, hierarchies, and typologies: Horses for courses. *Journal of Epidemiology and Community Health*, *57*, 527–529.

Poppelaars, M., Tak, Y. R., Lichtwarck-Aschoff, A., Engels, R. C., Lobel, A., Merry, S. N., Lucassen, M. F., & Granic, I. (2016). A randomized controlled trial

Appendices

Appendix A – Studies excluded from the review and rationale

Table 1

Studies excluded at full paper screening stage

Study	Exclusion criteria
Eichenberg, C., & Schott, M. (2017). Serious Games for Psychotherapy: A Systematic Review. *Games for*	

Study ID Number: 4

Type of Publication:
☐ Book/Monograph
☒ Journal Article
☐ Book Chapter
☐ Other (specify):

Domain:
☒ School- and community-based intervention programs for social and behavioural problems
☐ Academic intervention programs
☐ Family and parent intervention programs
☐ School-wide and classroom-based programs
☐ Comprehensive and coordinated school health services

General Design Characteristics

A1. Random assignment designs (if random assignment design, select one of the following)

☐ Completely randomized design
☒ Randomized block design (between participants, e.g., matched classrooms)
☐ Randomized block design (within participants)
☐ Randomized hierarchical design (nested treatments)

A2. Nonrandomized designs (if non-random assignment design, select one of the following)

☐ Nonrandomized design
☐ Nonrandomized block design (between participants)
☐ Nonrandomized block design (within participants)
☐ Nonrandomized hierarchical design
☐ Optional coding for Quasi-experimental designs

A3. Overall confidence of judgment on how participants were assigned (select one of the following)

☐ Very low (little basis)
☐ Low (guess)
☐ Moderate (weak inference)
☐ High (strong inference)
☒ Very high (explicitly stated)
☐ N/A
☐ Unknown/unable to code

B. Statistical Treatment/Data Analysis
☒ Appropriate unit of analysis
☐ Familywise/experimenter wise error rate controlled when applicable

C. Type of Program

☒ Universal prevention program
☐ Selective prevention program
☐ Targeted prevention program
☐ Intervention/Treatment
☐ Unknown

D. Stage of Program

☐ Model/demonstration programs

Study 4: Summary of Evidence

Indicator	Overall evidence rating 0-3	Description of evidence Strong Promising Weak No/limited evidence Or Descriptive ratings
General Characteristics		
Design		Pragmatic Randomised controlled trial
Type of programme		Universal prevention programme
Stage of programme		Early stage
Concurrent/ historical intervention exposure		

Table 4

WoE A Ratings

Study	Quality ratings assigned for the 4 dimensions			Overall WoE A (mean score to 2 decimal points)
	Measurement (0-3)	Comparison Group (0-3)	Sufficient Power (0-3)	

Weight of Evidence B

This criteria is derived from evidence typologies that suggest which designs are more appropriate for certain review questions (Petticrew & Roberts, 2003). For this review, WoE B considered the relevance of the methodology for evaluating the effectiveness of an intervention and the author selected design and generalisability as two key factors relating to effectiveness.

Table 5

WoE B Criteria

Criteria	Ratings	Rationale
Study Design	3.Definitive RCT 2. Pilot/Feasibility RCT 1.Non-randomised study (Cohort study, case-control studies, cross sectional survey, case reports)	Effectiveness relies on being able to conduct formal hypothesis testing for outcome measures and requires causality inferences to be made.

Generalizability of Study Sample	3. Multi-site, wide geographic coverage (e.g. across a country or multiple countries) 2. Multi-site, narrow geographic coverage (e.g. within a city) 1. Single-site (e.g. individual school)	Effectiveness relies on being able to generalize results to diverse student populations

Table 6

WoE B Ratings

Study	Quality ratings assigned for the 2 dimensions		Overall WoE B
	Study Design (1-3)	Generalizability of study sample (1-3)	
Perry et al., 2017	3	2	2.5
Poppelaars et al., 2016	3	3	3

Weight of Evidence C

This criteria is derived from Harden & Gough (2012) to discern topic relevance to the specific review question.

Table 7

WoE C Criteria

Criteria	Ratings	Rationale
Primary outcome measure of depressive symptoms	3. Primary outcomes include a valid, reliable and widely-used measure of depression symptoms for this specific population. 2. Primary outcomes include a non-validated or not widely used measure of depression for this specific population. 1. Depression is not included as a primary outcome (but is a secondary or exploratory outcome)	The review is concerned with depressive symptoms
Format of Delivery	3. Intervention delivered with flexibility (no structured monitoring or supervision of engagement) 2. A minimum threshold of engagement was enforced (some oversight and monitoring of engagement) 1. Highly structured and controlled oversight ensuring engagement	The aim of the review is to report on effectiveness rather than efficacy, therefore the most relevant studies are those where the intervention was delivered flexibly and compliance was not enforced ('real world' context / naturalistic)

Table 8

WoE C Ratings

Table 9

WoE D Ratings

Study	WoE A: Methodological Quality	WoE B: Methodological Relevance	WoE C: Topic Relevance	WoE D: Overall weight of evidence (average score of A,B and C)
Perry et al., 2017	High 3	High 2.5	High 2.5	High 2.67
Poppelaars et al., 2016	High 3	High 3	High 3	High 3
Kuosmanen et al., 2017	Medium 1.75	High 2.5	Medium 2	Medium 2.08
Fleming et al., 2012				