

***Case Study 1: An Evidence-Based Practice Review Report***

***Theme: School Based Interventions for Learning***

***How effective are utility-value interventions aimed at improving the academic performance of STEM students aged 18 to 25, in higher education settings?***

**Summary**

Utility value interventions (UVI)s ask students to write short essays or letters about the personal relevance of course material to their lives or the lives of others; helping students to make meaningful real life connections. UVIs are based on expectancy-value theory, which suggests that individuals pursue challenging tasks if they think they will succeed and if they find value in what they are learning (Eccles & Wigfield, 2002). This systematic literature review aims to examine how effective UVIs are at improving the academic performance of students aged 18 to 25, who are studying science, technology, engineering and maths (STEM) subjects in university. A systematic literature search was carried out on three databases and six studies met the inclusion criteria. One medium effect size was found, the rest were either small or negligible. As a result the findings suggest that UVIs may have minimal impact on the academic performance of young people aged 18 to 25 who are studying STEM subjects in higher education settings. Future research should seek to address the limitations that were identified and investigate the effectiveness of UVIs for underperforming students.

## **Introduction**

*What are utility value interventions?*

Utility value interventions (UVI)s target different psychological processes that are crucial for achievement. They are the perceived value of the course material and the engagement with course material (Brown, Smith, Thoman, Allen, & Muragishi, 2015). Researchers have employed various strategies for communicating utility value (UV) information. UVIs ask students to write about the relevance of academic content in the curriculum and how it may be relevant or useful to achieve personal goals (Hulleman, Durik, Schweigert, & Harackiewicz, 2008; Wigfield & Cambria, 2010). Another variation of a UVI entails evaluating UV quotations from other students (Gaspard et al., 2015). Becoming aware of the relevance of the curriculum and building connections helps students to appreciate the value of the course material, thus leading to a deeper level of engagement with the content and in turn improving performance (Klebanov, Burstein, Harackiewicz, Priniski, & Mulholland, 2017). For example, a student studying content about the skeletal system is likely to find greater value in the content if they are able to connect what they are learning to their future aspiration of becoming a physiotherapist. Whilst there are variations in the method, all interventions aimed at increasing student's self-perceived UV are considered as UVIs.

*Psychological relevance*

UVIs are based on Eccles' expectancy-value theory (Eccles & Wigfield, 2002) that posits it is an individual's expectations of success and subjective task value that determines their motivation, persistence and performance (Canning & Harackiewicz, 2015). To the extent that if individuals see value in a field of study and have the

belief that they will succeed in it, they are more likely to try to pursue that field of study. However, it may be easier to influence the subjective task value of students, rather than their expectations of success (Pajares, 1996). Eccles and Wigfield (1995) define utility as “the value a task acquires because it is instrumental in reaching a variety of long and short-range goals” (p.216). They believed that utility was ‘extrinsic’ as it can extend beyond the task at hand, as connections can be made with other tasks, activities or goals (Wigfield & Eccles, 1992). As a result UVIs attempt to promote perceptions of utility, as a means of facilitating learning, whilst also raising academic achievement, effort and interest (Hulleman et al., 2008; Wigfield & Cambria, 2010). Hulleman and Harackiewicz (2009) carried out a UVI that was embedded within science lessons. Over a semester, students were asked to write essays about how the curriculum was relevant to their lives. They found that it not only increased interest in science, but it also improved the academic performance of students who had low expectations of their success.

### *Rationale*

UVIs should be used to promote the pursuit of study of science, technology, engineering and maths (STEM) subjects at university level (Hecht et al., 2019). Increasing the supply of the next generation of STEM professionals has been a political priority across the world (Marginson, Tytler, Freeman & Roberts, 2013), particularly due to high attrition rates in STEM fields (Organisation for Economic Cooperation and Developments, 2015). The Green Paper recognised that there is a technical skills shortage within the UK, with the UK ranking 16 out of 20 OECD countries who have people with technical qualifications (HM Government, 2017). According to a Confederation of British Industry (CBI) /Pearson (2016) survey 40% of UK employers report difficulties recruiting staff due to a shortage of STEM

graduates. This failure to close the STEM skills gap has been attributed to a range of explanations: the perceived difficulties of STEM subjects, high dropout levels from sciences courses in higher education, low pay and poor career prospects (Seymour & Hewitt, 1997; Prados, 1998; Butz et al., 2006). In the US 48% of students who begin studying STEM majors in university change to a non-STEM field (Chen & Soldner, 2013). These stark figures highlight the importance of increasing interest and retaining university students who are studying STEM subjects worldwide.

### *Relevance to Educational Psychology*

According to research conducted by the Department for Education (2019) 26% of males and 34% of females want to pursue a STEM subject at A Level. This highlights that there is a desire to pursue STEM subjects. STEM subjects have high attrition rates, particularly at university so it is fundamental that EPs try to support universities, young people's wellbeing and academic performance to try and reduce this problem. Through doing so EPs will be making a valuable contribution to the life outcomes of young people. As a result of the Children and Families Act (2014) the role of Educational Psychologists (EPs) has been extended to cover up to 25 years of age (Department for Education, 2014). This is a significant change for EPs whom will need to develop practice and knowledge in order to meet the needs of the post-16 population (Atkinson, Dunsmuir, Lang & Wright, 2015). Baxter and Frederickson (2005) emphasise the need for EPs to widen their practice in order to promote the positive developmental outcomes for all children and young people not just those with special needs. Promoting the wellbeing and academic achievement of typically developing young people aged 18-25 in university is vital. If EPs increase their knowledge about the needs and required support of young people in university, their valuable contribution to higher education will be recognised. In addition, educational

psychology is considered a STEM profession, so by supporting students' pursuit of STEM subjects, EPs can support the future generation of the profession. To date EPs have been limited in how their role can be utilised within higher education settings in the UK. This review intends to inform EPs of UVIs that can support typically developing young people within higher education.

#### *Use of terminology*

Please note: In the USA higher education institutions are commonly called 'college', whereas in the UK they are called 'university'. This review will refer to 'university' to represent all higher education institutions. This review will also use The Higher Education Classification of Subjects (HECoS) listing of subject areas in the STEM, this outlines a wide range of subjects that are considered as STEM subjects, from biological sciences to psychology (Higher Education Statistics Agency, 2018).

#### *Review Question*

How effective are utility-value interventions aimed at improving the academic performance of STEM students aged 18 to 25, in higher education settings?

**Critical Review of the Evidence Base**

*Inclusion and Exclusion Criteria*

A comprehensive systematic literature search was carried out on 12/01/2020 using the electronic databases PsychINFO, Education Resources Information Centre (ERIC) and Medline. The search terms that were used are presented in Table 1. A filter was applied to include peer reviewed journals.

**Table 1**

*Search Terms*

1	2	3
utility value	intervention*	STEM
<b>AND</b>	<b>OR</b>	<b>OR</b>
	strateg*	biology
	<b>OR</b>	<b>OR</b>
	program*	chemistry
		<b>OR</b>
		physics
		<b>OR</b>
		science*
		<b>OR</b>
		technolog*
		<b>OR</b>
		engineer*
		<b>OR</b>
		math*

\*Denotes wildcard

The search yielded 91 results in total. 34 from PsychINFO, 40 from ERIC and 17 from Medline. 19 duplicates were removed and a further 62 journal articles were removed after the title and abstracts were screened according to the inclusion and exclusion shown in table 2. 10 journal articles were read in full and 4 were excluded with rationale (see Appendix A). The PRISMA flow chart (Figure 1) outlines the inclusion and exclusion process. 6 journal articles were selected for this review, these are summarised in Table 3 and are described in further detail in Appendix B.

**Table 2:**

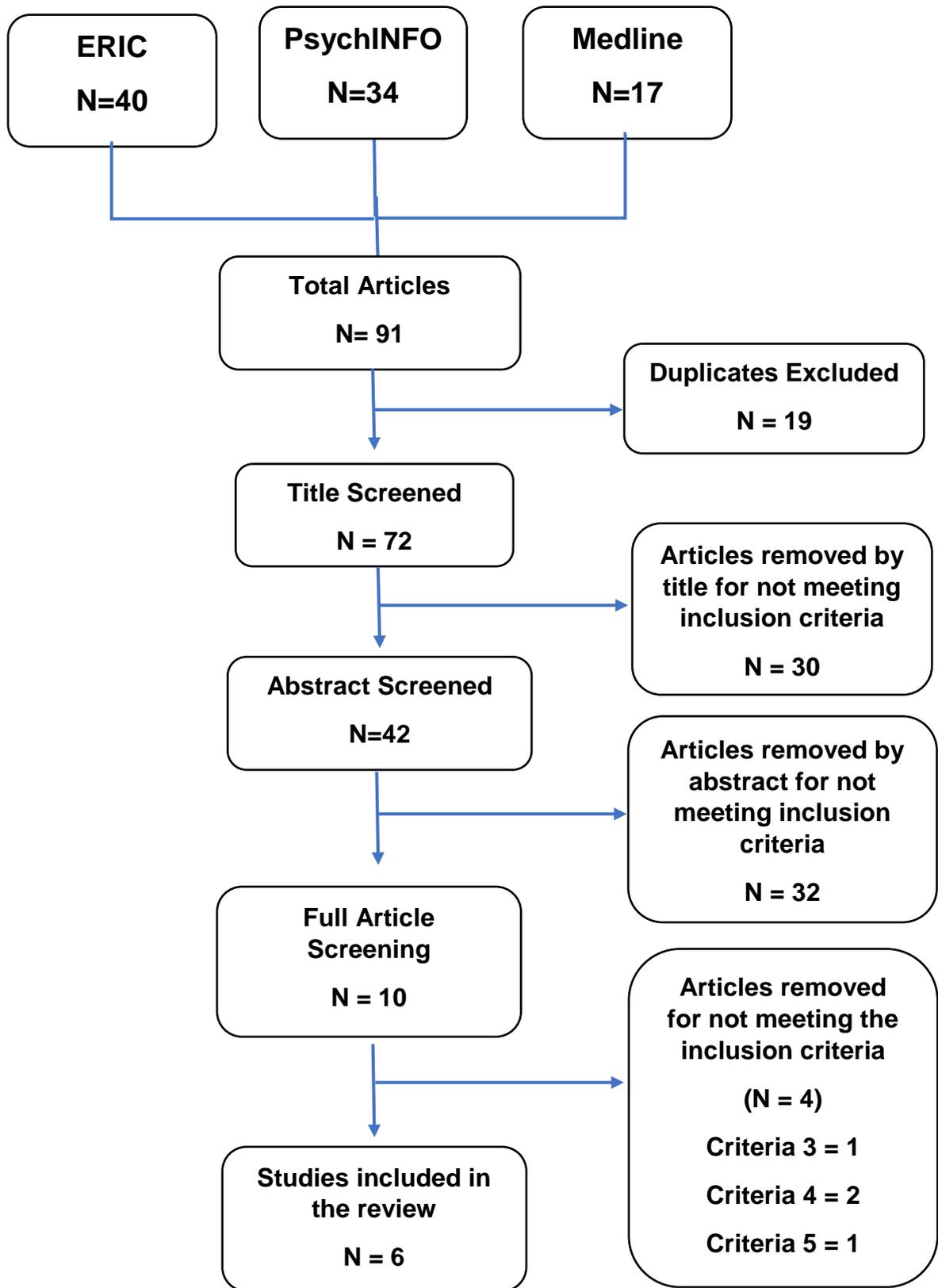
*Exclusion and Inclusion Criteria*

<b>Criteria</b>	<b>Inclusion Criteria</b>	<b>Exclusion Criteria</b>	<b>Rationale</b>
1. Type of publication	Articles from peer reviewed journals	Articles that are not peer reviewed	To ensure a minimum standard of methodological quality
2. Intervention	Intervention aims to increase self-perceived utility value of STEM subjects	Intervention does not aim to increase self-perceived utility value of STEM subjects	The purpose of the review is to critically evaluate utility value interventions of STEM subjects.
2. Language	Publications are produced in English	Publications that are not produced in English	Resources are not presently available to translate research studies
3. Setting	Intervention targets students who are studying STEM degrees that are taught at university	Intervention does not target students who are studying STEM degrees that are taught at university	To consider the effectiveness of UVIs for students who are studying STEM degrees in higher education
4. Design	Randomized Control or Group-based experimental design	Any other designs which is not group-based or publications which are not experimental e.g. reviews. Follow up studies are also excluded	To compare the effectiveness of utility value interventions between or within groups, Petticrew and Roberts, (2003) recommend using RCTs or group based experimental designs for effectiveness studies. Follow up studies are excluded as the focus of the review is not on developmental/longitudinal processes
5. Outcome variable	Study has outcome variable measuring	Study has outcome variable which does not measure	To assess the effectiveness of the utility value interventions in raising academic

<b>Criteria</b>	<b>Inclusion Criteria</b>	<b>Exclusion Criteria</b>	<b>Rationale</b>
	academic achievement	academic achievement	achievement as part of the review question

**Figure 1:**

*Flow Chart of Search Results:*



**Table 3:***Summary of Studies Included in Review*

---

<b>Included studies</b>
Canning, E. A., Harackiewicz, J. M., Priniski, S. J., Hecht, C. A., Tibbetts, Y., & Hyde, J. S. (2018). Improving performance and retention in introductory biology with a utility-value intervention. <i>Journal of Educational Psychology, 110</i> (6), 834–849.
Harackiewicz, J. M., Canning, E. A., Tibbetts, Y., Priniski, S. J., & Hyde, J. S. (2016). Closing achievement gaps with a utility-value intervention: Disentangling race and social class. <i>Journal of personality and social psychology, 111</i> (5), 745.
Hulleman, C. S., Kosovich, J. J., Barron, K. E., & Daniel, D. B. (2017). Making connections: Replicating and extending the utility value intervention in the classroom. <i>Journal of Educational Psychology, 109</i> (3), 387–404.
Priniski, S. J., Rosenzweig, E. Q., Canning, E. A., Hecht, C. A., Tibbetts, Y., Hyde, J. S., & Harackiewicz, J. M. (2019). The benefits of combining value for the self and others in utility-value interventions. <i>Journal of Educational Psychology, 111</i> (8), 1478–1497.
Rosenzweig, E. Q., Harackiewicz, J. M., Priniski, S. J., Hecht, C. A., Canning, E. A., Tibbetts, Y., & Hyde, J. S. (2019). Choose your own intervention: Using choice to enhance the effectiveness of a utility-value intervention. <i>Motivation Science, 5</i> (3), 269–276.
Rosenzweig, E. Q., Wigfield, A., & Hulleman, C. S. (2020). More useful or not so bad? Examining the effects of utility value and cost reduction interventions in college physics. <i>Journal of Educational Psychology, 112</i> (1), 166–182.

---

**Quality and Relevance of the studies**

Gough's (2007) Weight of Evidence (WoE) framework was used to evaluate the quality and the relevance of the studies included in the review. The framework outlines three dimensions which in total create an overall judgement of the suitability and quality of the studies. WoE A judgements consider the quality of the methodology, WoE B judgements consider the appropriateness of the methodology for answering the research question and WoE C judgements consider the appropriateness of the studies that have been included in answering the review question. These judgements were combined and averaged to create an overall assessment, which is WoE D. The combined WoE data is shown in Table 4, a further breakdown of the WoE criteria can be found in Appendix C. The Gersten et al. (2005) coding protocol was used to critically appraise the methodological quality of the studies that were included, as this was deemed the most fitting.

**Table 4**

*Weight of Evidence*

<b>Research</b>	<b>WoE A</b>	<b>WoE B</b>	<b>WoE C</b>	<b>WoE D</b>
Canning et al. (2018)	3 (high)	3 (high)	2 (medium)	2.7 (high)
Harackiewicz, Canning, Tibbetts, Priniski & Hyde (2016)	3 (high)	3 (high)	2 (medium)	2.7 (high)
Hulleman, Kosovich, Barron & Daniel (2017)	3 (high)	3 (high)	2.3 (medium)	2.8 (high)
Priniski et al. (2019)	3 (high)	3 (high)	2.5 (high)	2.8 (high)
Rosenzweig et al. (2019)	1 (low)	3 (high)	2.3 (medium)	2.1 (medium)
Rosenzweig, Wigfield, & Hulleman (2020)	3 (high)	3 (high)	2.5 (high)	2.8 (high)
<b>Please note: &lt;1.4 (low), 1.5-2.4 (medium) and &gt;2.4 (high)</b>				

*Participants*

All research studies were conducted in the United States of America (USA). As a result they were all given a medium rather than high WoE C rating for ‘Setting’, whilst the USA has similar cultural norms to the UK and students in both countries may experience similar university stressors, the findings may not be completely generalizable to the UK context. In total there were 3,118 participants across the 6 studies. Three of the studies reported the mean age of the participants (Hulleman et al., 2007; Priniski et al., 2019 and Rosenzweig et al., 2019), these ranged from 18.7 to 19.61. Priniski et al. (2019) and Rosenzweig et al. (2019) received a high rating on

WoE C, under 'Age of participants' as they also included the standard deviations for age, these ranged from 0.76 to 1.16. From this we can infer that all of the participants were close to this age range, this is important as the focus of the review was for young people between 18 to 25. Harackiewicz et al. (2016) and Canning et al. (2018) did not include the ages of participants so they received a low rating for the 'Age of participants' element of WoE C, however their overall WoE C was still 'medium'. Rosenzweig et al. (2020) received a medium WoE C rating as they included first to fourth year university students, although they did not include the specific ages of participants, from this we can conclude that the students were most likely to be between 18 to 25 as students who are enrolled in college on a 4 year course in the USA are typically of this age range (Bustamante, 2019). All of the studies included information about the ethnic make-up of participants, however not all studies included an ethnically diverse range of participants, this affected the WoE C rating that they were given as it may not be generalizable to a wide range of evidence. Canning et al. (2018), Rosenzweig et al. (2019) and Hulleman et al. (2017) received a medium WoE C rating as they did not include participants from any underrepresented minority (URM) students. Harackiewicz et al. (2016) received a medium WoE C rating as the focus of the study was only URM students. Priniski et al. (2019) and Rosenzweig et al. (2019) both included students from a range of different ethnic backgrounds and thus they were given a high WoE C rating.

### *Design*

All of the studies used a randomized pre-test post-test design. All of the participants were randomly assigned to either an experimental group or to a control group. Thus allowing us to infer that any observed differences between the students in the interventions group and the control group were due to the intervention, rather than participant variables. Randomised control trials are considered the gold standard

research design for evaluating treatments (Cochrane Collaboration, 2011). As a result all of the studies received high WoE B ratings.

### *Intervention*

All of the studies varied, either due to the population that they used, when they were carried out during the university school year or the procedure that was used.

Harackiewicz et al. (2016) investigated whether self-focused UVI can close the achievement gap for first generation (FG) and URM students. Students in the UV condition were asked to write self-focused essays. Priniski et al. (2019) investigated whether self-focused UVIs, other-focused UVI, a combination of both or a choice of UVI was more effective. Canning et al. (2018) investigated whether the dosage of UV assignments (0, 1, 2 or 3) as well as the timing of the UVI (at the beginning of the semester or the end of the semester) had an effect. Rosenzweig et al. (2019) investigated whether the choice of format of the UVI had an impact. Students were told either to write a letter or an essay, they were given a mixture of both or they were given the option to choose which format they wanted to use. Rosenzweig et al. (2020) investigated whether a self-focused UVI would be more effective than a cost-reduction intervention (CRI). Hulleman et al. (2017) compared an enhanced UVI which was designed to increase the frequency of connections made between the students' lives to a traditional self-focused UVI. All of the studies included meet the definition of UV originally given by Eccles and Wigfield (1995). As a result they received a high WoE C ratings under the 'level of description' heading unless, the procedures were not detailed enough. Rosenzweig et al. (2019), received a medium WoE C rating as no information was given about the number of participants in each condition and examples of the intervention instructions were not provided. Similarly, Priniski et al. (2019) also received a medium WoE C rating as examples of the instructions were not provided.

### *Fidelity of Implementation*

Fidelity of implementation is fundamental in order to prevent potentially false conclusions being drawn (Carroll, Patterson, Wood, 2007), about the UVI's effectiveness. All of the studies except Rosenzweig et al. (2019) attempted to monitor the fidelity of implementation, so they all received high WoE A ratings. Canning et al (2018) and Harackiewicz et al. (2016) both state that biology graduate students were hired to grade the UV assignments and to assess whether the students followed instructions. The other three studies assessed intervention fidelity in even greater depth. Hulleman et al. (2017) trained independent raters to assess writing quality, personalisation of connections and implementation intention. Priniski et al. (2019) assessed response frequency (assignment completion rates) and response quality (level of UV that is articulated). Rosenzweig et al. (2019) assessed engagement with task, time spent reading, ranking and evaluating the quotations.

### *Measures*

All of the studies collected pre and post intervention data, this contributed to the high WoE B rating that they received. Pre and post measures are important as they allow us to infer that any change in academic performance is as a result of the UVI. All of the studies reported the measures that were used to collect the data (Appendix B). Hulleman et al. (2017) and Rosenzweig et al. (2020) both used class exams scores as pre and post intervention measures for academic performance. Canning et al (2018), Priniski et al. (2019), Rosenzweig et al. (2019) and Harackiewicz et al. (2016) used prior grade-point average scores (GPA) as pre-intervention measures of academic performance and they used biology course grades as post intervention measures of academic performance. GPA is calculated by averaging accumulated final grades over time. It is commonly used as a predictor of academic performance, thus the operational validity of GPA can be measured as the degree to which it

correlates with academic performance (Bacon & Bean, 2006). Bacon and Bean (2006) found that overall GPA is highly correlated with academic performance ( $r=.94$ ), suggesting that it is a valid measure of academic performance. Ideally pre and post measures would be the same, however as achievement is the dependant variable it is not unusual that GPA is used as the pre-measure. There is no other measure of academic performance that rivals GPA (Richardson, Abraham & Bond, 2012), thus making it a useful pre-intervention measure to use.

### *Outcomes*

All of the effect sizes are presented in Table 5, they were calculated using the Beta values, if they were not included in the publications. The effect sizes were interpreted using Cohens  $d$  (1988) criteria which states a small effect is (0.2), a medium effect is (0.5) and a large effect is (0.8). Hulleman et al. (2017) reported the effect size. They combined the results of two UVIs, there was a small effect on the final exam scores of students (0.30), and individual UVI scores were not reported. Harackiewicz et al. (2016) reported a significant result of the UVI on academic performance, they also included the effect size, however it was considered negligible (0.06). This suggests that the effect of UVI was too small to infer that it had an effect on course grades for all students. Harackiewicz et al. (2016) also reported the effect of the UVI on FG-URM students and found a medium effect (0.55), suggesting that UVI may be more effective for URM and FG students. Rosenzweig et al. (2020) reported that there was a positive effect of the UVI on final exam scores, however they only reported a combined effect size of the UVI and the cost reduction intervention. As a result the effectiveness cannot be established independently for the UVI. In addition, as their study was underpowered the findings must be approached with caution.

Canning et al. (2018) reported that there was a significant effect of the UVI on biology course grades, however effect sizes were not reported, but through calculation, they were found to be small (0.24). This suggests that there may have been a small effect of the UVI on course grades. Priniski et al. (2019) reported that there was no overall effect of the UVI on course grades. After calculation, the effects was also negligible (0.06). Priniski et al. (2019) also reported that there was a main effect of choice on course grades, with students who were given a high choice UVI performing better than students in the low choice UVIs. After calculation, it was also considered negligible (0.16). The effect size indicates that the difference between the high choice and low choice UTI conditions are too small to infer that being in a high choice UVI will have an effect on course grades, compared to being in a low choice UVI. Rosenzweig et al. (2019) reported that there were no significant effect of any of the UVIs on final grades, effect sizes were also not reported. They combined the results from the high choice and low choice UVIs. After calculation, this effect was negligible (-0.04). This suggests that the effect was too small to infer that the UVI had an effect on course grades. This was the only study that received a 'low' WoE A rating, part of the reason for this was because effect sizes were not reported. It may be due to the poor methodological quality of the study that no effect was found. This contributed to the 'medium' WoE D rating that the study received. The rest of the studies received 'high' WoE D ratings as they were considered to be of higher methodological quality, more methodologically relevant and appropriate for investigating the review question.

All of the studies received either 'high' or 'medium' overall WoE D ratings, however none of them appear to have achieved large effect sizes. Only Harackiewicz et al. (2016) reported a medium effect of the UVI on FG-URM students. The rest of the

studies either found no effect, a negligible effect or a small effect. Consequently we cannot infer that UVIs influence the academic performance of all STEM students in higher education.

**Table 5**

*Effect sizes of studies*

Research	Sample Size	Design	Outcome Measure (Academic Performance)	Post-test outcome			
				Effect Size (Cohen's <i>d</i> )	Effect size interpretation	<i>p</i>	Weight of Evidence D
Canning et al. (2018)	577	Randomised Control Trial	Final biology course grades	0.24	Small effect	0.009	High
Harackiewicz et al. (2016)	1040	Randomised Control Trial	Final biology course grades for all students combined	0.06	Negligible effect	0.24	High
			Final biology course grades for URM and FG groups	0.55	Medium effect	*	
Hulleman, et al. (2017)	357	Randomised Control Trial	Class Exam Scores	0.25	Small effect	0.03	High

Research	Sample Size	Design	Outcome Measure (Academic Performance)	Post-test outcome			Weight of Evidence D
				Effect Size (Cohen's <i>d</i> )	Effect size interpretation	<i>p</i>	
Priniski et al. (2019)	590	Randomised Control Trial	Final biology course grades	(All UV conditions combined)	(most effective for low performing students)	0.68	High
				0.03	Negligible effect		
				(All UV conditions combined)	Negligible effect	0.40	
				0.16 (High choice UVIs vs low choice UVIs)			
Rosenzweig et al. (2019)	406	Randomised Control Trial	Final biology course grades	-0.04 (Overall UVI effect of high and	Negligible effect	>0.05	Medium

Research	Sample Size	Design	Outcome Measure (Academic Performance)	Post-test outcome			Weight of Evidence D
				Effect Size (Cohen's <i>d</i> )	Effect size interpretation	<i>p</i>	
Rosenzweig et al. (2020)	148	Randomised Control Trial	Average score from two physics exams. Exam 2 after the first session of the intervention and Exam 3 after the third session of the intervention.	0.30 (UVI and cost reduction intervention combined effect size)  low choice conditions)	Small effect  (most effective with low performing students)	0.01	High

## **Conclusion and Recommendations**

The aim of the systematic review was to assess the effectiveness of utility value interventions that were aimed at improving the academic performance of 18 to 25 year old students, studying STEM subjects in higher education. Little evidence was found to support the effectiveness of UVIs with only two studies having a small effect (Canning et al., 2019; Hulleman et al., 2017) and one study with a medium effect (Harackiewicz et al., 2016). As previously noted, the study by Rosenzweig et al. (2020) combined the results of the UVI with a CRI, so the effect size for this study was not considered, although they did report significant results. There was a medium effect of the UVI for FG-URM groups in one study (Harackiewicz et al., 2016). This group had the lowest grades prior to entering university and highest levels of high school poverty. This may explain why the study by Rosenzweig et al. (2019) was not effective and the study by Canning et al (2018) and Hulleman et al. (2017) only had small effect sizes, as the samples used did not include underrepresented groups. This is a major the limitation of all of these studies as the lack of diversity means they cannot be generalised to the general population.

### *Strengths and limitations*

Most of the studies received high WoE D ratings, except for Rosenzweig et al. (2019) which received a medium WoE rating, suggesting that the research overall was of high quality and relevant to the review question. Power calculations identified that the majority of the studies had good samples sizes, except for Rosenzweig et al. (2020) which was underpowered. In addition most of the studies reported that the findings were both significant and effective, so it was expected that there would be greater effects, however upon further exploration of the effect sizes, they were either small or negligible.

All of the UVIs were fully integrated into the course curriculums and they were issued as assignments which were completed outside of the university sessions. This does increase the ecological validity of the research, however it also raises questions about fidelity. This is because students in the different conditions may have communicated about the specific conditions that they were in, therefore influencing the answers that were given, or the level of impact of the UVI. It may be helpful to get the students to complete the assignments, in class and under controlled conditions in order to increase the validity of the findings. Another explanation for the small/negligible effect sizes may have been because students in the control groups were asked to summarise the material that they were using. Dunlosky, Rawson, Marsh, Nathan and Willingham (2013) found that summarising course material enhances learning. The control group task was essentially a cognitive intervention that also raises achievement. Future research should seek to use a more inert comparison group. All of the studies were conducted in the USA, they were conducted by common researchers and reportedly conducted in large or medium universities. This suggests that the studies may lack generalizability to the UK and smaller sized universities.

Strengths of the UVIs were that the course instructors were blind to the condition that the students were in, they were inexpensive to administer as they were issued online, the online delivery meant that students were randomized with ease and the data was easily available and accessible. In addition, UVIs can be easily adapted so that they can be used with different STEM subjects, the studies in this review alone, looked at UVIs used with biology, physics and psychology students.

*Recommendations for future research*

The studies by Hulleman, et al. (2017) and Rosenzweig et al. (2020) had small effects however they both reported that the UVI was most effective for low performing students. Similarly the FG-URM students in the study by Harackiewicz et al. (2016) were the lowest performing group and the UVI had a medium effect on their academic performance. This suggests that UVI may be more effective for low performing students. As a result EPs may be able to use UVIs to support low performing students in higher education, if the methodological issues are addressed. It would also be helpful to investigate the role of UVI for underperforming groups in secondary settings, as this has not been previously investigated. This would extend previous research by Harackiewicz and Hulleman (2009) who found that a UVI was more effective with students in high school who had low performance expectations. Future research should ensure that good sample sizes are used, diverse populations are included, control group tasks do not aid course performance, UVIs are investigated in a range of higher education settings, as well as the UK and UVI assignments are completed in controlled settings. EPs have a duty to support young people up to age of 25, so research must continue in order to understand which interventions, strategies and techniques are most effective for the post 16 population.

## References

- Atkinson, C., Dunsmuir, S., Lang, J., & Wright, S. (2015). Developing a competency framework for the initial training of educational psychologists working with young people aged 16–25. *Educational Psychology in Practice*, 31(2), 159–173.
- Bacon, D. R., & Bean, B. (2006). GPA in research studies: An invaluable but neglected opportunity. *Journal of Marketing Education*, 28(1), 35-42.
- Baxter, J. & Frederickson, N. (2005). Every child matters: Can educational psychology contribute to radical reform?, *Educational Psychology in Practice*, 21(2), 87-102.
- Brown, E. R., Smith, J. L., Thoman, D. B., Allen, J. M., & Muragishi, G. (2015). From bench to bedside: A communal utility value intervention to enhance students' biomedical science motivation. *Journal of Educational Psychology*, 107(4), 1116.
- Bustamante, J. (2019). *College Enrollment & Student Demographic Statistics*. Retrieved February 01, 2020, from <https://educationdata.org/college-enrollment-statistics/>
- Butz, W. P., Bloom, G. A., Gross, M. E., Kelly, T. K., Kofner, A., & Rippen, H. E. (2003). Is there a shortage of scientists and engineers? How would we know. *Santa Monica, Calif.: RAND Corporation, IP-241-OSTP*.

Canning, E. A., & Harackiewicz, J. M. (2015). Teach it, don't preach it: The differential effects of directly-communicated and self-generated utility–value information. *Motivation Science*, 1 (1), 47-71.

Canning, E. A., Harackiewicz, J. M., Priniski, S. J., Hecht, C. A., Tibbetts, Y., & Hyde, J. S. (2018). Improving performance and retention in introductory biology with a utility-value intervention. *Journal of Educational Psychology*, 110(6), 834.

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation science*, 2(1), 40.

Chen, X. (2013). STEM Attrition: College Students' Paths into and out of STEM Fields. Statistical Analysis Report. NCES 2014-001. *National Center for Education Statistics*.

Cochrane Collaboration (2011). *Cochrane handbook for systematic review of interventions*. Retrieved January 20, 2020, from <http://handbook.cochrane.org>

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Academic Press.

Confederation of British Industry. (2016). *The right combination: CBI/Pearson Education and Skills Survey 2016*. Retrieved January 20, 2020, from <https://epale.ec.europa.eu/sites/default/files/cbi-education-and-skills-survey2016.pdf>

Department for Education (DfE). (2014). *The Children and Families Act*. London: DfE.

Department for Education (DfE). (2019). *Attitudes towards STEM subjects by gender at KS4*, Social Science in Government. Reference: DFE-RB891. Retrieved January 21, 2020, from [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/777458/Attitudes\\_towards\\_STEM\\_subjects\\_by\\_gender\\_at\\_KS4.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/777458/Attitudes_towards_STEM_subjects_by_gender_at_KS4.pdf)

Didier, T., Kreiter, C., Buri, R., & Solow, C. (2006). Investigating the utility of a GPA institutional adjustment index. *Advances in Health Science Education*, 11, 145–153.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14, 4–58.

Eccles, J., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectations, values and academic behaviors. In J. T. Spence (Ed.), *Perspective on achievement and achievement motivation* (pp. 75–146). San Francisco, CA: Freeman.

Eccles, J. S., & Wigfield, A. (1995). In the mind of the achiever: The structure of adolescents' academic achievement related-beliefs and self-perceptions. *Personality and Social Psychology Bulletin*, 21(3), 215-225.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual review of psychology*, 53(1), 109-132.

Gaspard, H., Dicke, A. L., Flunger, B., Brisson, B. M., Häfner, I., Nagengast, B., & Trautwein, U. (2015). Fostering adolescents' value beliefs for mathematics with a relevance intervention in the classroom. *Developmental psychology*, 51(9), 1226.

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional children*, 71(2), 149-164.

Gough, D. (2007). Weight of evidence: a framework for the appraisal of the quality and relevance of evidence. *Research papers in education*, 22(2), 213-228.

Harackiewicz, J. M., Canning, E. A., Tibbetts, Y., Priniski, S. J., & Hyde, J. S. (2016). Closing achievement gaps with a utility-value intervention: Disentangling race and social class. *Journal of personality and social psychology*, 111(5), 745.

Hecht, C. A., Harackiewicz, J. M., Priniski, S. J., Canning, E. A., Tibbetts, Y., & Hyde, J. S. (2019). Promoting Persistence in the Biological and Medical Sciences: An Expectancy-Value Approach to Intervention. *Journal of Educational Psychology*, 111(8), 1462–1477.

Higher Education Statistics Agency (2018). *Proposed STEM Status of HECoS Subjects*. Retrieved January 21, 2020, from

[http://www.sfc.ac.uk/web/FILES/guidance\\_sfcgd212019/HECOS\\_STEM\\_Definition.pdf](http://www.sfc.ac.uk/web/FILES/guidance_sfcgd212019/HECOS_STEM_Definition.pdf).

HM Government (2017). *Building Our Industrial Strategy*, Green Paper. London: HM Government.

Hulleman, C. S., Durik, A. M., Schweigert, S. B., & Harackiewicz, J. M. (2008). Task values, achievement goals, and interest: An integrative analysis. *Journal of educational psychology*, 100(2), 398.

Hulleman, C. S., & Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. *science*, 326(5958), 1410-1412.

Hulleman, C. S., Kosovich, J. J., Barron, K. E., & Daniel, D. B. (2017). Making connections: Replicating and extending the utility value intervention in the classroom. *Journal of Educational Psychology*, 109(3), 387–404.

Klebanov, B. B., Burstein, J., Harackiewicz, J. M., Priniski, S. J., & Mulholland, M. (2017). Reflective Writing About the Utility Value of Science as a Tool for Increasing STEM Motivation and Retention—Can AI Help Scale Up?. *International Journal of Artificial Intelligence in Education*, 27(4), 791-818.

Marginson, S., Tytler, R., Freeman, B., & Roberts, K. (2013). STEM: Country comparisons: International comparisons of science, technology, engineering and

mathematics (STEM) education. Melbourne, Victoria: Australian Council of Learned Academies.

Organisation for Economic Cooperation and Development. (2015). *Education policy outlook 2015: Making reforms happen*. Retrieved 20 January, 2020, from <http://www.oecd.org/edu/education-policy-outlook-2015-9789264225442-en.htm>

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of educational research, 66*(4), 543-578.

Petticrew, M., & Roberts, H. (2003). Evidence, hierarchies, and typologies: Horses for courses. *Journal of Epidemiology and Community Health, 57*(7), 527–529.

Prados, J. W. (1998). *Engineering education in the United States: past, present and future*. Retrieved 18 January, 2020, from [www.ineer.org/Events/ICEE1998/lcee/papers/255.pdf](http://www.ineer.org/Events/ICEE1998/lcee/papers/255.pdf)

Priniski, S. J., Rosenzweig, E. Q., Canning, E. A., Hecht, C. A., Tibbetts, Y., Hyde, J. S., & Harackiewicz, J. M. (2019). The benefits of combining value for the self and others in utility-value interventions. *Journal of Educational Psychology, 111*(8), 1478–1497.

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: a systematic review and meta-analysis.

*Psychological bulletin*, 138(2), 353.

Rosenzweig, E. Q., Harackiewicz, J. M., Priniski, S. J., Hecht, C. A., Canning, E. A., Tibbetts, Y., & Hyde, J. S. (2019). Choose your own intervention: Using choice to enhance the effectiveness of a utility-value intervention. *Motivation Science*, 5(3), 269–276.

Rosenzweig, E. Q., Wigfield, A., & Hulleman, C. S. (2020). More useful or not so bad? Examining the effects of utility value and cost reduction interventions in college physics. *Journal of Educational Psychology*, 112(1), 166–182.

Seymour, E. and Hewitt, N. M. (1997). *Talking about leaving: why undergraduates leave the sciences*, Boulder, CO: Westview Press.

Wigfield, A., & Eccles, J. S. (1992). The development of achievement task values: A theoretical analysis. *Developmental Review*, 12, 1–46.

Wigfield, A., & Cambria, J. (2010). Students' achievement values, goal orientations, and interest: Definitions, development, and relations to achievement outcomes. *Developmental Review*, 30, 1–35.

## Appendices

### Appendix A: Excluded Studies at the Full Article Reading Stage, with Rationale

Excluded Research	
Reference	Rationale for Exclusion
Rozek, C. S., Hyde, J. S., Svoboda, R. C., Hulleman, C. S., & Harackiewicz, J. M. (2015). Gender differences in the effects of a utility-value intervention to help parents motivate adolescents in mathematics and science. <i>Journal of Educational Psychology</i> , 107(1), 195–206.	Intervention does not target STEM degrees which are taught at university
Beigman Klebanov, B., Burstein, J., Harackiewicz, J. M., Priniski, S. J., & Mulholland, M. (2017). Reflective writing about the utility value of science as a tool for increasing STEM motivation and retention- Can AI help scale up? <i>International Journal of Artificial Intelligence in Education</i> , 27(4), 791–818.	Any other designs which is not group-based or publications which are not experimental e.g. reviews. Follow up studies are also excluded
Hecht, C. A., Harackiewicz, J. M., Priniski, S. J., Canning, E. A., Tibbetts, Y., & Hyde, J. S. (2019). Promoting Persistence in the Biological and Medical Sciences: An Expectancy-Value Approach to Intervention. <i>Journal of Educational Psychology</i> , 111(8), 1462–1477.	Any other designs which is not group-based or publications which are not experimental e.g. reviews. Follow up studies are also excluded
Brown, E. R., Smith, J. L., Thoman, D. B., Allen, J. M., & Muragishi, G. (2015). From bench to bedside: A communal utility value intervention to enhance students' biomedical science motivation. <i>Journal of Educational Psychology</i> , 107(4), 1116–1135.	Study has outcome variable which does not measure academic achievement

Appendix B: Mapping the Field

Authors	Participants/Setting	Intervention	Design	Pre-intervention measures	Post-intervention measures
<b>Canning et al. (2018)</b>	577 students who were enrolled on an introductory biology course at a large university in the US	There were 3 sessions, with a 5 week gap between each. Students in the UV condition wrote about the relevance of course material to their lives or the lives of others. Students in control condition wrote an essay related to content that was covered in a lecture. The number and timing of the UVI assignments varied.	RCT with eight conditions testing dosage (0, 1, 2 or 3 UV) and timing of the UVI (early in the semester).	GPA	Final biology course grades
<b>Harackiewicz, et al. (2016)</b>	1040 underrepresented minority ethnic and/or first generation students who were enrolled on an introductory biology course at a large US university	Students in the UVI completed three UV assignments across the semester which asked them to discuss the relevance of the course material to their own or others' lives. Students in the control condition summarised course material. Students in the value affirmation (VA) condition wrote about personal values that were important to them and students in the VA control	RCT with 4 conditions (UV, UV control group, VA and VA control group)	GPA	Final biology grades

Authors	Participants/Setting	Intervention	Design	Pre-intervention measures	Post-intervention measures
<b>Hulleman et al. (2017)</b>	357 students on an introductory psychology class at a mid-sized university in the US. 84% were white. The mean age of participants was 18.7 years	condition were asked to write about values that were not important to them.  Three writing assignments over the course of a semester. Students in the control condition were asked to summarise course material. Students in the other two conditions were asked to relate what they had learnt to their lives. Students in the EUV condition had to complete a further 3 prompts that were aimed at increasing the UV of their subject area, this was in addition to completing the same task as the UV condition. They also completed questions that prompted reflection of the implementation strategies that they discussed from intervention one.	RCT with 3 conditions: enhanced UVI, UVI and control group	Class Exam Score	Class Exam Score
<b>Priniski et al. (2019)</b>	590 students who were enrolled on an introductory biology course at a large university in the US. The mean age was 19.61	3 writing assignments over the course of a semester. The control group summarised course material. Students in the other five conditions completed UV assignments relating either to themselves or/and others depending on the condition that they were in.	RCT with 6 conditions: control, UVIs -self-focused, other focused choice between self or group –focused, self and other focused UVI in that order and the other	GPA	Final course grades

Authors	Participants/Setting	Intervention	Design	Pre-intervention measures	Post-intervention measures
<b>Rosenzweig et al. (2019)</b>	406 students with a mean age of 19.19 who were enrolled on an introductory biology course at a large U.S. university. First-generation college students and underrepresented racial/ethnic minority students, were not include	The UVI comprised of three writing assignments over the course of a semester. Students in the high-choice UVI condition had the option of choosing essay or letter formats for each of the 3 writing assignments. Students in the two low-choice UVI conditions were asked to complete either an essay and then a letter or a letter and then an essay. They were able to choose the format for the third assignment. Students in the control condition were given course material to summarise.	and self-focused UVI  RCT with four conditions: high UVI, two different low UVI conditions or a control group	GPA	Course grades
<b>Rosenzweig et al. (2020)</b>	148 students who were enrolled on an introductory physics course at a large mid-Atlantic U.S. university	The UVI comprised of three writing assignments over the course of a semester. Students in the UVI and CRI conditions read and evaluated quotations from students and wrote quotations in response. The UVI quotations related course material to student’s lives. The CRI quotations described challenges students faced in physics classes and they were overcome. There were two control groups. One control	RCT with three conditions: UVI, cost reduction intervention or control group	1 <sup>st</sup> Course Exam Grades	Average score from two exams. Exam 2 after the first session and Exam 3 after the third session.

---

Authors	Participants/Setting	Intervention	Design	Pre-intervention measures	Post-intervention measures
		group did not do anything and the other was a summary condition.			

---

## Appendix C – Weight of Evidence (WoE)

### ***WoE A: Methodological quality***

The Gersten et al. (2005) coding protocol was used to critically appraise the methodological quality of the studies that were included. Each study was rated for both essential and desirable quality indicators. The table below illustrates the cut off points that were used. One question on the protocol was changed from “Was sufficient information provided to determine/confirm whether the participants demonstrated the disability (ies) or difficulties presented?” to “Was the sample population appropriate for the research hypothesis/hypotheses?”. The rationale for this was that the focus of the review is typically developing children and not children with special educational needs.

### **Weighting criteria for WoE A:**

WoE A Rating	Criteria
<b>3</b> <b>(High)</b>	<ol style="list-style-type: none"> <li>1. The study meets at least 9 essential criteria</li> <li>2. The study meets four or more desirable criteria</li> </ol>
<b>2</b> <b>(Medium)</b>	<ol style="list-style-type: none"> <li>1. The study meets at least 9 essential criteria</li> <li>2. The study meets at least 1 and less than 4 desirable criteria</li> </ol>
<b>1</b> <b>(Low)</b>	<ol style="list-style-type: none"> <li>1. The study meets less than 9 essential criteria</li> </ol>

Based on recommendations from Gersten et al. (2005)

**Summary of Weight of Evidence (WoE A) Ratings**

<b>Research</b>	<b>Essential Criteria</b>	<b>Desirable Criteria</b>	<b>WoE A Rating</b>
Canning et al. (2018)	9	5	3 (High)
Harackiewicz, et al. (2016)	10	4	3 (High)
Hulleman et al. (2017)	10	6	3 (High)
Priniski et al. (2019)	9	5	3 (High)
Rosenzweig et al. (2019)	7	3	1 (Low)
Rosenzweig et al. (2020)	10	8	3 (High)

**WoE B: Methodological Relevance to the review question**

The methodological relevance of the studies considered how appropriate the design of each study was for answering the review question.

High (3)	Medium (2)	Low (1)
<b>Randomised control trials</b>	<b>Quasi-experimental and Cohort designs</b>	<b>Non-experimental designs</b>
-Pre-test/post-test data collection	-Pre-test/post-test data collection	-Single-variable research
-Control/comparison group	-Control/comparison group	-Correlational
		-Quasi-experimental studies (with no control/comparison group)
		-Qualitative studies

The criteria is based on recommendations from Petticrew and Roberts (2003) for research that is best suited for studying the 'effectiveness' of interventions

*Summary of WoE B ratings:*

Research	WoE B Rating
Canning et al. (2018)	3 (High)
Harackiewicz, et al. (2016)	3 (High)
Hulleman et al. (2017)	3 (High)
Priniski et al. (2019)	3 (High)
Rosenzweig et al. (2019)	3 (High)
Rosenzweig et al. (2020)	3 (High)

**WoE C: Relevance to the review question**

The table below outlines, the criteria, breakdown of ratings and rationale for WoE C:

Criteria	WoE C Rating	Rationale
A. Level of Description	<p>3. The procedure of the UVI is described in detail at each stage, including examples of intervention instructions                      - The UVI is based on Eccles and Wigfield’s (1995) definition of UV</p> <p>2. The procedure of the UVI is described at each stage, not including examples of intervention instructions                      - The UVI is based on Eccles Wigfield’s (1995) definition of UV</p> <p>1. The procedure of the UVI is not described in enough detail to allow for replication                       The UVI is not based on Eccles definition of UV</p>	Detailed procedure allows for replicability of the research.
B. Age of Participants	<p>3. The age range of participants is included and it is between 18-25, the mean age is included with standard deviations</p> <p>2. The mean age of participants is included and it is between 18-25. The year that the students are in university is included.</p> <p>1. The age of participants is not included</p>	The review aims to understand the effectiveness of UVIs for students who are 18-25 as little is known about how we can support this age group, therefore it is valuable to know the age ranges that are included
C. Ethnicity of Participants	<p>3. The study will include participants from a diverse range of ethnic backgrounds (more than 3)</p> <p>2. The study will not include participant from a diverse range of ethnic backgrounds</p> <p>1. The study will not include information about the ethnic background of participants</p>	The reviews aims to include interventions that are applicable to students from a range of different ethnic backgrounds

Criteria	WoE C Rating	Rationale
D. Setting	3. The research will be conducted in a UK university 2. The research will be conducted in a university in a country that has similar cultural norms to the UK 1. The research will be conducted in a country that does not have similar cultural norms to the UK	To increase the generalisability of the findings, as participants in the UK or those from countries with similar cultural norms are likely to have shared experiences

**Summary of WoE C ratings:**

Research	A	B	C	E	WoE C Rating
Canning et al. (2018)	3	1	2	2	2
Harackiewicz et al. (2016)	3	1	2	2	2
Hulleman et al. (2017)	3	2	2	2	2.3
Priniski et al. (2019)	2	3	3	2	2.5
Rosenzweig et al. (2019)	2	3	2	2	2.3
Rosenzweig et al. (2020)	3	2	3	2	2.5

**WoE D: Overall Weight of Evidence Judgment**

An overall evidence rating for WoE D was established by calculating the mean value of WoE A, B and C. The scores of each WoE rating is illustrated below:

<b>Research</b>	<b>WoE A</b>	<b>WoE B</b>	<b>WoE C</b>	<b>WoE D</b>
Canning et al. (2018)	<b>3 (high)</b>	<b>3 (high)</b>	<b>2 (medium)</b>	<b>2.7 (high)</b>
Harackiewicz, et al. (2016)	<b>3 (high)</b>	<b>3 (high)</b>	<b>2 (medium)</b>	<b>2.7 (high)</b>
Hulleman et al. (2017)	<b>3 (high)</b>	<b>3 (high)</b>	<b>2.3 (medium)</b>	<b>2.8 (high)</b>
Priniski et al. (2019)	<b>3 (high)</b>	<b>3 (high)</b>	<b>2.5 (high)</b>	<b>2.8 (high)</b>
Rosenzweig et al. (2019)	<b>1 (low)</b>	<b>3 (high)</b>	<b>2.3 (medium)</b>	<b>2.1 (medium)</b>
Rosenzweig et al. (2020)	<b>3 (high)</b>	<b>3 (high)</b>	<b>2.5 (high)</b>	<b>2.8 (high)</b>

**Please note: <1.4 (low), 1.5-2.4 (medium) and >2.4 (high)**

## Appendix D – Coding Protocol

**Coding protocol:** Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C, & Innocenti, M. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children*, 71,149-164.

**Study:** Rosenzweig, E. Q., Wigfield, A., & Hulleman, C. S. (2020). More useful or not so bad? Examining the effects of utility value and cost reduction interventions in college physics. *Journal of Educational Psychology*, 112(1), 166–182.

### ***Essential Quality Indicators - Quality indicators for describing participants***

Was the sample population appropriate for the research hypothesis/hypotheses?

Yes

No

N/A -

Unknown/Unable to Code

Were appropriate procedures used to increase the likelihood that relevant characteristics of participants in the sample were comparable across conditions?

Yes

No

N/A

Unknown/Unable to Code

Was sufficient information given characterizing the interventionists or teachers provided? Did it indicate whether they were comparable across conditions?

Yes

No

N/A

Unknown/Unable to Code

### ***Essential Quality Indicators - Quality indicators for Implementation of the Intervention and Description of Comparison Conditions***

Was the intervention clearly described and specified?

Yes

No

N/A

Unknown/Unable to Code

Was the fidelity of implementation described and assessed?

Yes - Level of engagement, time spent reading, ranking and evaluating the quotations was assessed

No

N/A

Unknown/Unable to Code

Was the nature of services provided in comparison conditions described?

Yes

No

N/A

Unknown/Unable to Code

***Essential Quality Indicators – Quality Indicators for Outcome Measures***

Were multiple measures used to provide an appropriate balance between measures closely aligned with the intervention and measures of generalised performance?

Yes

No

N/A

Unknown/Unable to Code

Were outcomes for capturing the intervention's effect measured at the appropriate times?

Yes

No

N/A

Unknown/Unable to Code

**Essential Quality Indicators – Quality Indicators for Data Analysis**

Were the data analysis techniques appropriately linked to key research questions and hypotheses? Were they appropriately linked to the limit of analysis in the study?

Yes

No

N/A

Unknown/Unable to Code

Did the research report include not only inferential statistics but also effect size calculations?

Yes

No

N/A

Unknown/Unable to Code

**Desirable Quality Indicators**

Was data available on attrition rates among intervention samples? Was severe overall attrition documented? If so, is attrition comparable across samples? Is overall attrition less than 30%?

Yes

No

N/A

Unknown/Unable to Code

Did the study provide not only internal consistency reliability but also test-retest reliability and interrater reliability (when appropriate) for outcome measures? Were data collectors and/or scorers blind to study conditions and equally (un)familiar to examinees across study conditions?

Yes

No

N/A

Unknown/Unable to Code

Were outcomes for capturing the intervention's effect measured beyond an immediate post-test?

Yes

No

N/A

Unknown/Unable to Code

Was evidence of the criterion-related validity and construct validity of the measures provided?

Yes

No

N/A

Unknown/Unable to Code

Did the research team assess not only surface features of fidelity implementation (e.g. number of minutes allocated to the intervention or teacher/interventionist following procedures specified), but also examine quality of implementation?

Yes

No

N/A

Unknown/Unable to Code

Was any documentation of the nature of instruction or series provided in comparison conditions?

Yes

No

N/A

Unknown/Unable to Code

Did the research report include actual audio or videotape excerpts that capture the nature of the intervention?

Yes

No – Not for all of the studies

N/A

Unknown/Unable to Code

Were results presented in a clear, coherent fashion?

Yes

No

N/A

Unknown/Unable to Code

**Essential Quality Indicators Total Score: 10/10**

**Desriable Quality Indicators Total Score: 8/8**

**Quality rating - 3**