Evaluation and Program Planning 83 (2020) 101857

Contents lists available at ScienceDirect

Evaluation and Program Planning

journal homepage: www.elsevier.com/locate/evalprogplan

Development of a fostering purpose intervention[★,★★,★]

Brian R. Riches[a,*], Celina M. Benavides[b], Valeska X. Dubon[a]

[a] Claremont Graduate University, United States
[b] California State University, Los Angeles, United States

ARTICLE INFO

Keywords:
Purpose in life
Evolutionary evaluation
Positive psychology intervention
Adolescents
Young adults
Open educational resource

ABSTRACT

Evolutionary Evaluation is a framework for understanding the evolution of programs. Just as programs grow and develop, so can standalone interventions. This paper focused on building and evaluating a purpose intervention to scaffold adolescents and young adults' search for purpose. Four studies were conducted at each phase of the intervention's lifespan to assess if modifications were needed and to test the viable validity of using online activities to increase both the search for and the discovery of a purpose in life. These studies can be viewed through the lens of the Evolutionary Evaluation framework. Study 1 evaluated the intervention at its initiation phase, which consisted of developing and testing the intervention activities. Study 2 tested the intervention during the development phase to determine which activities most effectively increased rates of purpose. Study 3 evaluated the intervention in its stability phase by experimentally testing all activities together in a full curriculum via Amazon's Mechanical Turk. Finally, study 4 assessed the dissemination of the intervention by testing the full curriculum with over 500 high school and college students. These studies suggest that the fostering purpose intervention cultivated purpose among adolescents and young adults. Additionally, the success of the intervention demonstrates that the Evolutionary Evaluation framework may be an effective model with which to develop an intervention, from initiation through dissemination phases.

1. Development of a fostering purpose intervention

Purpose is an intrinsically motivating desire to contribute to the world beyond-the-self, by engaging with meaningful, far horizon goals (Damon, Menon, & Bronk, 2003). Young people with purpose are healthier, demonstrate greater well-being, and are more successful in school than their peers without purpose (Benson, 2006; Pizzolato, Brown, & Kanny, 2011). A few programs attempted to foster purpose in adolescents and young adults, but due to their format and structure, they are limited in their ability to reach large numbers of young people. To address this limitation, an online fostering purpose intervention was developed. The intervention scaffolds adolescents' and young adults' search for purpose by providing opportunities to reflect on the things that matters most to them. The current paper describes the design process and provides evidence for the efficacy of the fostering purpose curriculum. Our curriculum development can be interpreted through the lens of the Evolutionary Evaluation framework.

1.1. Importance of fostering purpose

Purpose in life is a developmental asset that is linked to numerous positive outcomes, including life satisfaction, mental well-being, and academic success (Benson, 2006; Bundick, Yeager, King, & Damon, 2010). Unfortunately, purpose in young people is rare. Only 20 % of high school students and about 30 % of college students report a purpose in life (Damon, 2008). Parents, community practitioners, and educators are increasingly interested in fostering purpose in adolescents and young adults through high quality and convenient interventions (Picciano & Seaman, 2007).

Efforts to cultivate purpose are worthwhile given the rarity and importance of purpose among young people. However, there is limited research examining the ways that purpose can be intentionally fostered

[★] Note, the Fostering Purpose Toolkit discussed in this paper can be accessed at https://www.fosterpurpose.org/.

[★★] Note, We would like to thank Gabe Sanchez and Joe Hurst for their support and input during this project, and offering their classroom time for building student purpose over many years.

[★] Note, Some data from study 2 was previously published in Bronk, K. C., Baumsteiger, R., Mangan, S., Riches, B., Dubon, V., Benavides, C., & Bono, G. (2019). Fostering purpose among young adults: Effective online interventions. *Journal of Character Education*, *15*(2), 21-38.

[*] Corresponding author at: Division of Behavioral and Organizational Sciences, Claremont Graduate University, ACB 124, 150 E. 10th Street, Claremont, CA, 91711, United States.
E-mail address: brian.riches@cgu.edu (B.R. Riches).

https://doi.org/10.1016/j.evalprogplan.2020.101857
Received 4 October 2019; Received in revised form 10 July 2020; Accepted 30 July 2020
Available online 05 October 2020
0149-7189/

in the lives of adolescents and young adults. To date, purpose-fostering interventions have been tried in educational, therapeutic, and work settings (Dik, Steger, Gibson, & Peisner, 2011; Frankl, 1984; Pizzolato et al., 2011). One unintentional intervention discovered that college students who participated in a 45-minute guided interview about the things that matter most to them demonstrated higher scores on purpose and life satisfaction nine months after the interview (Bundick, 2011). This finding suggests that purpose can be cultivated when there are scaffolded opportunities for young people to reflect on and discuss their goals and core values. However, a one-on-one discussion with each student is not a feasible way of reaching large numbers of young people. Instead, drawing on this approach, our intervention translated the interview into a set of online tools that could be administered to large numbers of young people to relatively quickly and easily foster purpose (Bronk et al., 2019). The development of our fostering purpose intervention can be interpreted in light of the recommendations of Evolutionary Evaluation.

1.2. Evolutionary evaluation

Just as species evolve over time, so too do programs. Program variations that are effective survive, while variations that do not work in particular contexts are discarded. Evolutionary Evaluation (EE) is an approach, or overarching framework, for evaluating programs. EE has informed the development of specific evaluation protocols (e.g., The Systems Evaluation Protocol) and has been used to evaluate character education programs (Urban, Linver, Thompson, Davidson, & Lorimer, 2017). EE is a useful framework to draw upon given the focus on methodologically appropriate evaluation at each phase of the intervention in its lifecycle. EE suggests specific procedures for each of the four phases of program development: initiation, development, stability, and dissemination.

1.2.1. Similarities between intervention phase and evaluation framework

According to EE, alignment is critical in the creation of an intervention. Alignment entails matching an appropriate evaluation approach to the intervention's stage in the intervention life cycle (Urban, Hargraves, & Trochim, 2014). In other words, interventions should be evaluated differently based on the lifecycle phase of the program. The four intervention phases (i.e., initiation, development, stability, and dissemination) should be aligned to corresponding evaluation phases that include process and response, change, comparison and control, and generalizability (Urban et al., 2014). The alignment of the intervention and evaluation phases ensures that information gained through the evaluation is appropriate given the intervention's life cycle (Urban et al., 2014). Additionally, EE (Urban et al., 2014) recommends the methods that should be used to evaluate a program at each point in its life cycle. For example, while qualitative methods can be used at any evaluation phase, they are especially beneficial in the initiation phase to evaluate the process and response to the intervention and allow for quick changes after initial trials (Urban et al., 2014).

The goal of this study was to develop a purpose intervention. Throughout the development and testing of this intervention, we kept in mind the evaluation process (i.e., planning, implementation, and improvement; Patton, 1987). We created a logic model to better understand the inputs and outputs of the intervention and used formative and summative evaluation to examine fidelity of implementation, and the effectiveness of the intervention, while also making improvements where necessary. While not aligned exactly, our process can be interpreted in light of the EE framework (Urban et al., 2014). Although Urban et al.s' (2014) study was published when we were already in the process of our evaluation, after examining the EE framework closely, we retrospectively discovered that the steps we took were similar to EE. As such, we report on the design and efficacy of the fostering purpose curriculum and how our process can be interpreted through the lens of EE.

One distinction is important as EE is often used to examine the "natural selection" of programs and program components across their lifespan, but also recognizes that programs can be consciously modified or selected through an "artificial selection" process (Urban et al., 2014). Urban et al. (2014) note that "both natural and artificial selection follow the same evolutionary rules of variation and selective retention (p. 130)." The decisions made in the phases of our project served as more of a systematic "artificial selection" process for identifying program components that increased purpose. We combined these "artificially selected" activities into a full intervention. Our approach is similar to the EE framework, but offers a novel use of EE in artificially selecting purpose fostering activities, rather than selecting entire program components through the intervention's evolution.

2. Study 1: initiation evaluation

In the initiation phase, an intervention is new and undergoing rapid change. Given its early developmental stage, the EE framework suggests a process and response evaluation using qualitative methods. This entails focusing on the intervention implementation and participant and facilitator satisfaction (Urban et al., 2014). In our project, we began a qualitative pilot study that asked for participant feedback of our activities and measures to ensure understanding and engagement.

The pilot was conducted with high school students who completed activities where they reflected on their best possible selves, analyzed quotes about purpose, and ranked their values from most important to least important. Quantitative measures showed increases in purpose, prosocial intentions, and academic aspirations. Open ended, qualitative responses allowed the researchers to assess how the activities were received and how they could be improved. Additionally, students and their teachers participated in a focus group where they discussed the intervention material and provided feedback. Based on this early feedback, the intervention was modified.

2.1. Participants

A total of 356 adolescents were recruited from a high school in southern California ($Mage$ = 16.90, SD = .44). Most participants (51.8 %) identified as female (46.5 % identified as male, and 1.3 % identified as another gender). Additionally, 36.2 % of participants identified as White, 35.5 % as Latino, 12.7 % identified as mixed race/ethnicity, 6.4 % as Asian, 3.3 % Black, and 1.8 % Native Hawaiian or other Pacific Islander. Students were compensated $5 USD. Adolescents were included in the study if they had parental consent or were 18 years of age or older and provided their own consent, and if they passed attention checks in the surveys.

2.2. Procedure

At the beginning of the school year, teachers were trained to facilitate the intervention. Fourteen classrooms were randomly assigned to treatment or control conditions. Five classrooms served as controls: two were Advanced Placement (AP) classes (64 students) and three were non-AP classes (72 students). Control groups completed memory skill activities, such as learning mnemonics, that were facilitated by the teachers.

Students in the three treatment conditions participated in activities for one week. Treatment Group "Quotes" included four classrooms (87 students) and analyzed quotes, watched videos, and responded to related writing prompts. Treatment Group "Best Possible Selves (BPS)" contained three classrooms (72 students) and completed best possible selves activities, values sorting, and related writing prompts. Finally, the third treatment Group "AP All" included two AP classes (61 students) that were given all the intervention activities that the Quotes and BPS groups received.

Students completed control or intervention activities during class

time. The activities took approximately 40 min a day, or one class period, for five days. Students completed quantitative pre and posttests before and after the week of activities. Students also responded to open-ended questions each day where they reported on their level of engagement, gave feedback on the day's activities, and reported on their teacher's level of engagement. Additionally, teachers responded to open-ended questions each day where they reported on the delivery of the intervention, their level of engagement, and how the activities could be improved. In addition, classrooms were randomly observed for implementation quality and teacher engagement.

2.3. Measures

2.3.1. Purpose in life

Purpose in life was measured using an early version of the Claremont Purpose Scale (CPS; Bronk, Riches, & Mangan, 2018). This version of the CPS consisted of 15 items measuring each of the three components of purpose (i.e., meaning in life, goal-orientation, and beyond-the-self). Sample items included: "I have a good sense of what makes my life meaningful" and "I have long-term goals I am working toward." Participants responded to these items on a five-point Likert type scale. Cronbach's alphas for this study ranged from .93 to .98.

2.4. Materials

2.4.1. Fostering purpose activities

There were two primary activities designed to foster purpose. The first involved reading and responding to purpose quotes, asking students what purpose entails, reflecting on their own purpose, and sharing what the quotes taught them about working to achieve that purpose. In the second activity, best possible selves, students imagined themselves in the future and reflected on short- and long-term goals that could help them become the best possible version of themselves. Other activities included a Values Q sort designed to help students reflect on what personal values were most important to them, videos to explain purpose in life, and a tattoo activity to help students visualize the things that mattered most to them. These activities were delivered through physical handouts and PowerPoint presentations facilitated by teachers.

2.4.2. Control activities

The control activities focused on teaching memorization strategies. Participants spent approximately 40 min each day on these activities. These activities were delivered by teachers in the same manner as the treatment activities.

2.5. Quantitative results

No significant differences in age, gender, or ethnicity were found between groups. Paired samples t-tests revealed non-significant decreases on the CPS from the pretest ($Mpre = 5.67$, $SD = .84$) to posttest ($Mpost = 5.65$, $SD = .83$) ($t(63) = .26$, $p = .795$) in the AP control group, as well as the non-AP control group ($Mpre = 5.30$, $SD = .95$) ($Mpost = 5.29$, $SD = .98$) ($t(71) = .17$, $p = .866$). Additionally, the Quotes ($Mpre = 5.42$, $SD = .98$) ($Mpost = 5.27$, $SD = 1.06$) ($t(86) = 1.71$, $p = .091$) and BPS ($Mpre = 5.49$, $SD = .73$) ($Mpost = 5.41$, $SD = .86$) ($t(71) = 1.00$, $p = .319$) groups showed non-significant decreases in purpose from pretest to posttest. The only group to increase on purpose in life ($Mpre = 5.34$, $SD = .93$) ($Mpost = 5.50$, $SD = 1.03$) ($t(60) = -1.59$, $p = .117$) was the AP All group which included two classes that were given all the intervention activities. This evidence suggests that the fostering purpose activities worked best when combined, rather than split apart.

To further compare the difference between groups, we calculated change scores on the CPS from pre to posttest and made comparisons using independent samples t-tests. While the difference between the treatment and control group in the AP classes was the most different in the sample, the difference was not statistically significant ($t(123) = -1.46$, $p = .147$). Neither of the non-AP treatment groups were significantly different from the non-AP control group [Quotes group ($t(157) = -1.22$, $p = .224$), BPS group ($t(142) = -0.69$, $p = .491$)].

We were curious as to why the intervention groups did not increase in purpose when there were statistically significant increases during the pilot testing. We hypothesized that differences in activity implementation might be the cause of a discrepancy, and as such, we examined the classroom observation notes and qualitative surveys from teachers and students.

2.6. Qualitative results

The qualitative results showed that students generally enjoyed the activities and were engaged. Several students commented: "I had lots of fun doing this activity!" and "It was really eye-opening and really made me think about a purpose in my life". However, there were also critical comments, such as, "What if people have no idea what they want to do with life? What do they do then?" These criticisms helped to pinpoint which activities might be confusing or required further revisions. It was also beneficial to learn that there were a few universally loved activities, including one project that asked students to create a meaningful tattoo. Overall, student opinions were positive. There were no universally disliked activities, instead many activities were enjoyed by one full classroom of students, but not enjoyed by another classroom of students. To explore this, teacher implementation was examined using teacher reports and teacher observations.

Teacher implementation quality varied across activities, rather than systematically across classrooms or teachers. This might explain the lower than expected increase in quantitative purpose scores. Other than the universally loved tattoo activity, individual teachers ranked some activities very positively or very negatively. Each teacher appeared to enjoy and engage in some activities more than others, leading to variable implementation quality across all teachers.

Classroom observations provided further insight to implementation differences. When teachers were not engaged in the delivery of the intervention this coincided with teachers reporting that they did not enjoy the activities for that class period. For example, this was apparent in an observation of a teacher who was facilitating different intervention activities during two different class periods on the same day. In the earlier period, the observer reported that the teacher was prepared, attentive, began the activities on time, the students were engaged, and the teacher facilitated interactive discussions. Conversely, in a later period the same teacher deviated from the activity and led an off-topic discussion for long enough that the observer believed they had mistakenly attended the wrong class. When the teacher returned to the topic, they did not engage with students and the class discussion was minimal. By the end of the period, the teacher lamented that they found this activity difficult to deliver, and that was why the students were off task. It appears that when the teacher enjoys the activities, the students were more engaged. The classroom observations and feedback helped guide intervention improvements.

Some teachers implementing the intervention delivered the majority of the curriculum in an online format. At the request of these teachers, a few classrooms completed the intervention during class time on laptops or mobile devices, with minimal teacher participation. In classroom observations, these students were more on task and better able to transition to classroom discussions in comparison to classrooms with full teacher implementation. This led us to consider implementing the activities in a fully online format. An online format could also address other problems that arose during implementation, including teachers' difficulties using videos or PowerPoints given varying technological skill levels. Additionally, if the activities were fully online, students might be able to complete them without teacher interaction, providing a more consistent implementation quality. For example, on

some days there were students who were absent. Given the limited time available to complete interventions in the public school system, these students were unable to make up the activities. This was not tracked nor planned, which made it impossible to account for this variable in quantitative statistical analyses. If these activities were fully online, students who missed the activities in class could have the opportunity to complete them later at home. Implementing the intervention online could increase the consistency in delivering the activities and reach a greater number of students.

2.7. Discussion

The qualitative and quantitative feedback during the initiation phase of the intervention contributed to modifying the intervention activities and delivery method. EE framework recommends using qualitative feedback as a primary source of information in this phase and doing so contributed to developing and changing the intervention as needed. The daily qualitative surveys offered insight to which activities teachers and students enjoyed and engaged with more deeply. The qualitative recommendations also helped identify common problems that could be solved through immediate and future modifications. Using quantitative experimental methods at this stage of the intervention's evolution is discouraged under the EE framework. In this initiation evaluation, purpose increased only in the group that completed all intervention activities, not in groups who completed only a specific subset of activities. These results illustrated one of the reasons why Randomized Control Trials (RCT) are discouraged, and represent misalignment at the initiation phase in the EE framework. A RCT allowed an assessment of the activities in their current state, but qualitative feedback showed implementation variability and a dislike of some activities, among other contextual factors that indicated the activities needed to be modified. In other words, the RCT assessed a set of rapidly changing activities and if we focused solely on the unfavorable quantitative results we would have presumed that the activities did not work, when instead the qualitative data offered insight to changes that should be made to the intervention.

2.7.1. Viable validity

Viable validity is an important concern at each intervention phase under the EE framework (Urban et al., 2014). Questions about viable validity pertain to the input of relevant stakeholders (e.g., students, teachers) on whether an intervention is affordable, practical, useful, helpful in the real world, and if it can be implemented without assistance from the authors or evaluating research team (Chen, 2010; Urban et al., 2014). Viable validity also evaluates if an intervention can recruit or retain participants outside of research contexts where participants are compensated, a program would also need to be evaluable (Chen, 2010). While each aspect of viable validity may not be tested in each of the current studies it is important to consider the wider scope of a program being viable in the real world at each stage of a program's lifespan. In this case the primary stakeholders are school teachers, guidance counselors, and the young people themselves. Our intervention demonstrated some viable validity by being affordable, practical for teachers or students to administer without assistance from the research team, and useful in solving relevant problems of student lack of purpose, to name just a few examples.

In study 1 viable validity was evaluated by observing classroom implementation and gathering qualitative feedback from the teachers. The primary problems to viable validity discovered through this phase were implementation and engagement variability. Most teachers wanted a more streamlined and uniform delivery of activities. Additionally, teachers who implemented the activities mostly or entirely online indicated more favorable opinions of the activities, as well as more participant engagement. These insights from stakeholders led us to move the activities to a fully online delivery system.

A fully online intervention might demonstrate more viable validity because it could increase the ease of use, practicality, and affordability of implementation without the assistance of the authors. Additionally, the intervention could be designed to be delivered during an average class period with little to no teacher involvement. This enables teachers to use the intervention activities individually or along with other curriculum or character education programs in a manner that is most convenient to their classroom, making it more practical for a public school setting.

By delivering fostering purpose activities in a uniform, online format, the likelihood of random implementation variation diminishes. Online learning tools allow flexible access by providing teachers, parents, and students access to content and instruction at any time or place (Means, Toyama, Murphy, Bakia, & Jones, 2009). Additionally, the intervention could be used outside of class time or assigned as homework so as not to overburden teachers whose instructional time is already stretched thin preparing students to meet state and federal academic standards. Online intervention activities could also be used by home schooled students, incorporated into larger programs, or used by individual young people who want to increase purpose in their lives outside of a formal structure. After these modifications, we moved to the development stage with a clear sense that the intervention should be fully online and accessible for free.

3. Study 2: development evaluation

The development phase of intervention evolution is characterized by "changes or revisions; however, the scale and scope of those revisions are smaller than what is seen during initiation" (Urban et al., 2014, p. 132). Dramatic changes are still possible during the development phase, but the intervention is becoming more and more stable as it is repeatedly tested. Urban et al. (2014) recommend that during this stage the focus should be on reliability and validity of measures, and participant outcomes in specific contexts, rather than the generalizability of the intervention. This may be done through pre and posttests of measures of interest.

During the next phase of the project, individual intervention activities were tested using a pre and posttest via Amazon's Mechanical Turk (MTurk), a very specific context, with a young adult sample. The goals were to discover which individual activities fostered purpose in young people, make improvements to the effective activities, and establish a stable set of activities that could be used together in a multi-day intervention.

3.1. Participants

Participants were recruited through MTurk, an online crowdsourcing platform enabling researchers to access a generally representative sample of the broader U.S. population. Research using MTurk samples points to increased sample diversity when compared to other convenience samples (Behrend, Sharek, Meade, & Wiebe, 2011; Buhrmester, Kwang, & Gosling, 2011; Summerville & Chartier, 2013). A total of 14 participants who failed attention check items were removed, leaving 565 participants included in the final sample (see Table 1 for sample size by group). The mean age for the sample was 24.61 (SD = 5.68) with an age range of 18–30 years old. There were 296 males and 265 females. The majority (75.6 %) identified as White, 9.7 % identified as Black, 6.0 % identified as Asian, 5.6 % were Latino and 3.1 % identified as another race/ethnicity. Most participants (85.7 %) competed some college credit. To qualify for the study, participants had to reside in the U.S., be 18 or over, and be fluent in English. Participants were compensated $2 USD for their time.

3.2. Procedure

Participants completed all study components using computers or mobile devices. Each participant was provided a pretest to measure

Table 1
Paired samples *t*-test results for all measures.

Condition (*n*)	CPS Full Scale		CPS Meaning		CPS Goals		CPS Beyond-the-Self		SPI	
	t	*M* Change	*t*	*M* Change	*t*	*M* Change	*t*	*M* Change	*t*	*M* Change
1 BPS & VIA-IS (48)	−0.98	−0.09	−0.51	−0.06	−1.45	−0.17	−0.35	−0.04	0.12	0.02
2 BPS & Tattoo (67)	0.83	0.06	2.34*	0.24	−0.06	0.01	−0.73	0.06	3.91***	0.47
3 BPS & Video (66)	2.83**	0.22	2.73**	0.34	1.56	0.14	2.25*	0.18	4.39***	0.52
4 BPS & Qsort (72)	2.76**	0.21	3.39**	0.42	2.11*	0.19	0.38	0.03	4.71***	0.70
5 Quotes & VIA-IS (54)	−0.87	−0.06	1.69	0.14	−1.22	0.12	−2.35*	0.19	1.62	0.24
6 Quotes & Tattoo (61)	1.74	0.12	2.64*	0.30	0.08	0.01	0.63	0.05	4.19***	0.52
7 Quotes & Video (61)	1.2	0.09	2.18*	0.34	−1.78	0.19	1.77	0.13	3.44**	0.42
8 Quotes & Qsort (71)	1.4	0.11	1.95	0.26	0.35	0.03	0.57	0.04	4.36***	0.64
Control (64)	−3.56**	−0.22	−2.83**	−0.22	−3.68***	−0.29	−1.90	−0.15	−8.80***	−2.17

Notes. Best Possible Self (BPS), Values in Action Inventory of Strengths (VIA-IS), Claremont Purpose Scale (CPS), Searching for Purpose Inventory (SPI).
* $p < .05$.
** $p < 0.01$.
*** $p < .001$.

purpose in life and searching for purpose. All items had a five-point Likert-type response ranging from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*). After the pretest, each participant was randomly assigned to one of nine conditions. Each condition included two activities that took approximately 10 min to complete. The control condition consisted of two memory activities. Each of the eight intervention conditions consisted of two purpose-enhancing activities presented in a counterbalanced order to eliminate order effects and isolate effects of each individual activity (see Table 1 for descriptions of conditions). After completing the activities, participants completed the posttest which included a few open ended questions to assess intervention delivery. Participants were asked, "What, if anything, did you learn about yourself from these activities?", as well as open ended questions to report any confusion, frustration, and positive feedback about the process of completing the activities. Finally, participants were compensated via MTurk.

3.3. Measures

3.3.1. Purpose in life

As noted previously, an early version of the CPS (Bronk et al., 2018) was used to examine the three components of purpose (i.e., meaning in life, goal-orientation, and beyond-the-self; $\alpha = .93$–$.94$). While the CPS is intended to measure identified purpose as an entire scale, many of the activities were designed to increase either meaning, goal-orientation, or beyond-the-self concerns so we also examined the subscales of the CPS.

3.3.2. Searching for purpose

Searching for purpose was measured using an early version of the Searching for Purpose Inventory (SPI) that contained six self-developed items. Participants were asked to think about the past three days and answer questions like, "I thought about what is most important to me in life" and "I thought about how I want to contribute to the world." The items measured active thinking about the three components of purpose in life. Cronbach's alpha ranged from .89 to .91. The SPI was used as an intention check as well as a measure of purpose to assess if participants were completing activities that should increase their purpose as well as measuring actual increases in their searching for purpose.

3.4. Materials

3.4.1. Fostering purpose toolkit

The fostering purpose intervention included online activities that were easy to access and disseminate through Qualtrics. Participants were asked to reflect on their values, goals, what was presently meaningful to them, and what effect they wished to have on the world beyond themselves.

3.4.2. Purpose introduction video

A nine-minute video summary of purpose was created by our team (Adolescent Moral Development Lab, 2017). The video described purpose in lay terms, explained its components, and gave examples to help participants understand purpose.

3.4.3. Quote activities

Participants read a quote about purpose in life and were asked to write an open-ended response about what the quote meant to them and how the quote related to what mattered most in their life.

3.4.4. Best-possible-self activity

These activities were a modified version of Layous, Nelson, and Lyubomirsky's (2013) Best Possible Self (BPS) activity. These activities asked participants to imagine themselves in the future, what they would be doing, what would be most important to them, and to set goals to get to their ideal future.

3.4.5. Celebrity video activities

Participants viewed videos of celebrities discussing their purpose. There were a range of different actors, entertainers, and sports stars of various genders and ethnicities that shared how they found their purpose and what was most meaningful to them in life.

3.4.6. Q sort activity

In the Q sort, participants ranked 13 statements of values, such as "supporting social issues is very important to me," "serving my family is very important for me," and "helping others is very important for me." These values were placed into one of three categories: "Exactly Like Me," "Neither Like Me or Not Like Me," and "Not at all Like Me." Participants were limited to only three values in the "Exactly Like Me" and "Not at all Like Me" categories to increase their self-awareness of what was most important to them.

3.4.7. Tattoo activity

Participants wrote about the type of person they are, the things that matter most to them, and their goals in life. Then, they were tasked to design a tattoo that would symbolize who they are and what they value. This activity was previously completed in school classrooms with students applying temporary tattoos to themselves. In this fully online format, participants created and described their tattoo design representing their purpose in life or picked an image on screen that best represented their purpose.

3.4.8. Values assessment

In this activity, participants took the Values in Action Inventory of Strengths (VIA-IS; Peterson, Park, & Seligman, 2005). Afterwards,

participants were given a brief report on what they scored highest on, allowing for reflection on individual values.

3.4.9. Control activities

The control activities focused on teaching memorization strategies.

3.5. Results

No significant differences were found between the treatment and control in regards to age, gender, ethnicity, or education level. Paired samples t-tests were performed for each of the conditions. The control group decreased significantly on both identified purpose ($t(63) = -3.56$, $p < .01$) and searching for purpose ($t(63) = -8.80$, $p < .001$), as well as the CPS subscales of meaning ($t(63) = -2.83$, $p < .01$) and goal orientation ($t(63) = -3.68$, $p < .001$), but did not change significantly on the beyond-the-self subscale.

All the intervention groups, with the exception of the two VIA activity groups, increased significantly on searching for purpose (see Table 1). Furthermore, participants in group 3 who completed the BPS and watched a video about purpose ($t(65) = 2.83$, $p < .01$) and group 4 who completed the BPS and a Q sort ($t(71) = 2.76$, $p < .01$) increased significantly on the full CPS scale. Group 2 who completed the BPS and created a tattoo ($t(66) = 2.34$, $p < .05$), group 6 who completed quote activities and created a tattoo ($t(60) = 2.64$, $p < .05$), and group 7 who read purpose quotes and watched the video ($t(60) = 2.18$, $p < .05$) increased significantly on the CPS meaning subscale. The VIA activities were not effective at increasing searching for, or identified, purpose. In fact, group 5 participants who read purpose quotes and completed the VIA decreased significantly on the CPS beyond-the-self subscale ($t(53) = -2.35$, $p < .05$). These results indicated that six of the seven online fostering purpose activities did increase searching for and identifying a purpose in life. For a summary of all analyses, please see Table 1.

3.6. Discussion

Viable validity was again assessed in this phase by examining participant responses in the activities, along with open ended questions about the delivery of the intervention. A few comments allowed us to fix technological errors with materials not appearing correctly, otherwise the delivery of the activities was received positively by the participants and the activities were presented as intended. The more relevant qualitative data was gained by examining participants' answers to activity questions. Length, relevance, and depth of participant responses to activity questions gave some indication of participant motivation. It appeared that all participants who answered survey attention check items correctly took the activities seriously. While there was a range of length and depth to responses, this was expected as each participant should be at a different stage in their search for purpose. We expected some participants to be at a developmental phase where they may not have searched for purpose at all, some participants may already have a clear sense of purpose, and most participants appeared to find and clarify their purpose though these activities. Due to this variety we did not exclude any participants whose qualitative responses were short or otherwise seemed to indicate lower motivation. While participants themselves are one relevant stakeholder this study did not assess the viability of the intervention with the other most relevant stakeholder, teachers. This study did not involve teachers in any way and while it only assesses a small aspect of the wide scope of viable validity the move to online activities should help increase the practicality of the intervention, as well as the affordability with all activities being offered for free.

In the Evolutionary Evaluation spirit of development and change, after discovering the success of these six activities, an additional two activities were added to the intervention. An email activity asked participants to message one to five adults they respected and knew well, such as mentors, coaches, close friends, and family members. Participants were provided an email template asking the recipient to reflect on the participants' strengths and talents, and how they could see the participant contributing to the world in the future. About five days later, participants who received responses were asked how they felt when reading their messages. This activity was designed to help participants reflect on their purpose, as well as gain input from a person the participant respected. A gratitude activity, based off the "three good things" activities, was also added to the purpose intervention (Seligman, Steen, Park, & Peterson, 2005). The exercise was designed to stimulate participants' thinking outside of themselves to help expand their capacity for beyond-the-self thinking. These activities were evaluated using the same methods described above and proved effective in increasing MTurk participants' purpose scores.

We also tested another implementation of the fostering purpose activities that tailored the activities a participant received based on their lowest score for the subscales on the CPS. For example, participants who scored low on the meaningfulness subscale of the CPS were presented with activities that were shown to increase personal meaningfulness, such as the BPS, tattoo, and quotes activities. While the activities did increase purpose in participants overall, they failed to increase the specific components of purpose the activities directed them towards (e.g., meaning or beyond-the-self concerns) significantly more than activities that were shown to increase other components of purpose (e.g., goal orientation). A possible reason for these results is that the early version of the CPS was not as reliable and did not provide enough valid data about each subscale to be used in this manner. This failure to tailor activities to a person who needs more personal meaningfulness, goal orientation, or beyond-the-self concerns to increase their purpose is a clear demonstration of the development phase of Evolutionary Evaluation. According to EE, the development phase of an intervention is marked by changes to the intervention based on evaluation results. During this phase, the intervention was still evolving and tested repeatedly as it became more stable. When the intervention demonstrated consistent stability over repeated testing, we moved forward in testing its stability as a full intervention with all the effective activities delivered to participants as a single curriculum.

4. Study 3: stability evaluation

One difficulty of using the EE lens to evaluate a program through its evolution is deciding when a program has moved to a lifecycle phase where it is stable as opposed to developing. Programs evolve over their lifespan and may need to move from a phase of stability to a phase of development after some time. This can make it difficult to determine if an evaluation should focus more on assessing the change of development or the comparison and control recommended in a more stable program. This problem is further exasperated when we examined our evaluations in light of the EE approach retrospectively.

Study 3 is the first time the individual activities were combined in a multi-day format where participants completed all activities. Additionally, a synthesis activity was added that could not be tested individually without completing all other activities beforehand. While this was the natural evolution of the program it suggested that the program in this form may again be in the initiation phase and may need to be altered further and a process and response or development evaluation might be most informative. However at this point in the program, the individual activity components were already studied, refined, selected, and stabilized. This suggested that future refinements were not likely to be as large and a stability evaluation would be most informative at this point in the program's evolution. Thus, we chose to focus our evaluation on the stability of the program while including aspects of the process and response evaluation similar to studies 1 and 2.

Evaluations at the stability phase of an intervention are characterized by a focus on comparison and control (Urban et al., 2014). Experimental and quasi-experimental methods can examine if the

intervention caused a change in purpose. In other words, we examined if the intervention demonstrated internal validity. While the terms comparison and control inspire thoughts of controlled experiments, evaluations of stable interventions can also incorporate comparative qualitative methods.

After discovering which activities were effective at fostering a purpose in life, those activities were combined into a full intervention. Adolescents and young adults completed two to three fostering purpose activities a day, over the course of four days. The full intervention was tested experimentally through Amazon's MTurk. We expected participants who completed this longer intervention would increase in both searching for purpose and identified purpose, while participants in the control condition would not.

4.1. Participants

A total of 169 participants were recruited through MTurk, with 73 in the intervention group and 96 in the control group. The mean age was 23.7 (SD = 4.11), with a range from 18 − 28. The sample was approximately even across males and females (51.3 % male). The majority (66.7 %) identified as White, 12.1 % identified as Black, 7.6 % identified as Asian, 7.6 % were Latino, and 4.5 % identified as another race/ethnicity. Most participants (90.6 %) indicated that they had completed at least some college credit. The same eligibility requirements for study 2 were used in this study.

4.2. Procedure

Participants completed a pretest to measure their purpose in life and searching for purpose, with the order of scale items randomized across participants. All items had a five-point Likert-type response ranging from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*). After the pretest, each participant was randomly assigned to one of two conditions. Each condition completed four sessions or days of activities; each session took approximately 30 min to complete. The treatment group completed fostering purpose activities while the control group completed memory activities. After completing the activities, participants took a posttest. Subsequent surveys were sent to participants, with a lagged follow up about one week later and again two weeks afterward. Participants were compensated $9.50 USD for their time.

4.3. Measures

Purpose in life was measured using the CPS (Bronk et al., 2018); the Cronbach's alpha ranged from .89 to .94. Searching for purpose was measured using the Searching for Purpose Inventory (Cronbach's alpha = .79–.88). Both measures were described in prior sections.

4.4. Materials

4.4.1. Fostering purpose toolkit

The purpose intervention included four online sessions that asked participants to reflect on their values, goals, what was presently meaningful to them, and what effect they wished to have on the world beyond themselves. Most of the activities were described in study 2, however one additional activity (detailed below) was included to synthesize the curriculum on the last day of the intervention.

4.4.2. Synthesis activity

This activity was designed to summarize and synthesize what participants experienced in the intervention. Participants were reminded about their reported future plans and ideal world. They were then asked to set specific goals for the next month, year, and five years to work towards their purpose.

4.4.3. Control activities

These activities focused on memorization strategies. Participants completed four sessions of activities, each 15 − 20 min in length. During the first session, participants practiced memorization skills; later activities built on these skills leading to other types of memory skills, such as location memory. Participants also learned the MAPS (Music, Association, Picturing, Stories) technique to remember individual pieces of information. These activities were delivered through Qualtrics in the same manner as the intervention activities.

4.5. Results and discussion

The focus of study 3 was to evaluate the internal validity of the fostering purpose intervention as a full and stable curriculum and examine the effects of the activities in the weeks after the intervention among late adolescents and young adults. However, through the lens of EE, the program at this phase could also be seen as being in the initiation or development phase. As such, additional studies were conducted that are not discussed in this paper. In one study we attempted to tailor activities to increase participants' purpose to specifically increase their beyond the self concerns, develop personal meaning, or increase goal directedness, depending on which component they scored lowest. While this did not successfully increase individual components of purpose it offered more focus on implementation, participant satisfaction, and qualitative assessments of change. Another study examined the stability of the program more explicitly without examining changes using a process and response evaluation (Bronk et al., 2019). These studies suggested that the stability evaluation used in study 3 was appropriate with our additional qualitative measure included.

No significant differences were found between intervention and control groups on age, gender, ethnicity, or education level. Paired samples t-tests were performed for each of the conditions. As expected, there was no significant difference from the pre to posttest on the CPS (t(62) = 1.50, $p > .05$) or the SPI (t(68) = -0.13, $p > .05$) for the control group.

Searching for purpose and identified purpose did increase in the intervention group. Paired samples t-tests revealed significant increases from pretest (M = 5.54, SD = .91) to posttest (M = 5.74, SD = .82) on the CPS (t(65) = 2.08, p = .041) and with the Searching for Purpose Inventory pretest (M = 5.55, SD = 1.05) to posttest (M = 6.06, SD = .78) (t(65) = 3.90, $p < .001$) for the intervention group. This evidence suggested that the fostering purpose activities helped adolescents and young adults search for purpose, which led to an increase in purpose in life over the short span of one week.

To further compare the difference between the control and treatment groups, we calculated change scores on the CPS and SPI from pre to posttest for both groups, and used independent samples t-tests to make group comparisons. We found that there were significant differences in pre to post changes between the control and intervention group on the CPS (t(120) = 2.39, p = .018) and SPI (t(133) = 3.05, p = .003). This showed that while the increase in purpose for the treatment group had a relatively small effect size, participants' change in purpose was significantly greater than it would have been without the fostering purpose activities.

Finally, we conducted longitudinal analyses to examine the change in purpose over time and how long the effect of the fostering purpose intervention lasted. We examined data across the four time points. Time 1 included the pretest survey that participants completed prior to commencing the four-day purpose intervention. Time 2 included the posttest that was provided immediately after the final day of activities, and the remaining time points consisted of the two lagged follow-ups that were each spaced one week apart.

Longitudinal data was available from 45 purpose intervention participants. We did not include control group participants in the longitudinal follow up surveys, given that there was no change in their purpose from the pretest to posttest. No significant differences were

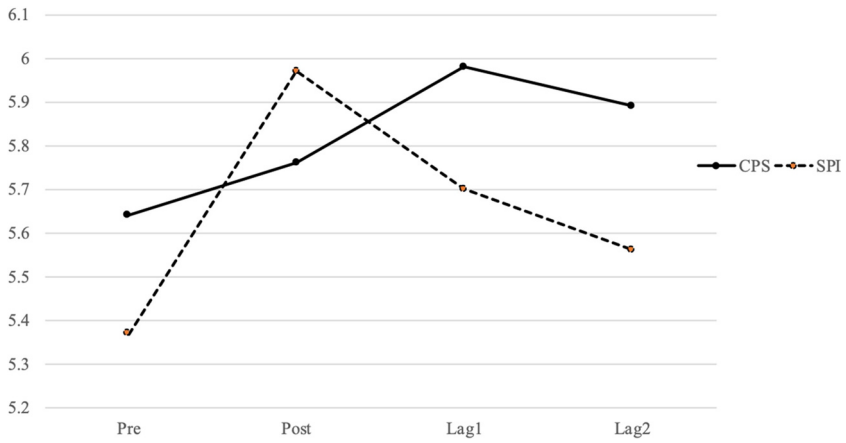

Fig. 1. CPS and SPItreatment group mean scores across time points.

found on age, gender, ethnicity, or education level for participants who completed all time points and those who did not. The treatment group's trend on purpose in life was compared using a General Linear Model. As shown in Fig. 1, the increase in purpose (measured by the CPS) depicted a significant linear trend ($F(1, 23) = 4.533$, $p = .044$) between the pretest ($M = 5.64$, $SD = 0.68$), posttest ($M = 5.76$, $SD = 0.80$), and lag tests 1 ($M = 5.98$, $SD = 0.68$) and 2 ($M = 5.89$, $SD = 0.78$). There was a moderately significant quadratic trend in searching for purpose ($F(1, 16) = 3.05$, $p = .100$), such that, the intervention group demonstrated a significant increase in searching for purpose from pretest ($M = 5.37$, $SD = 1.11$) to posttest ($M = 5.97$, $SD = 0.76$), and a decrease in searching for purpose after the intervention was completed at lag 1 ($M = 5.70$, $SD = 0.82$) and 2 ($M = 5.56$, $SD = 0.90$).

These results indicated that participants' purpose in life increased during the week of the intervention and the week following the intervention, and this increase in purpose remained relatively constant for a few weeks after the intervention. Additionally, there was a significant increase in searching for purpose in the intervention group during the week of the intervention. However, participants reported that they searched for purpose less and less at the lag 1 and 2 follow ups (Figs. 2 and 3). This suggested that the fostering purpose activities guided participants to search for their purpose over the course of the intervention activities and while some found more purpose during the week of the intervention, many participants continued to clarify their purpose in the week after the intervention concluded. Qualitative responses within the activities reinforced the quantitative results indicating that participants took the activities seriously, understood them, were motivated to complete all activities, and felt satisfied with the activities. Additionally, our qualitative measures showed that the implementation of the program was effective, and many participants showed clear development in their search for purpose over the course of the intervention.

5. Study 4: dissemination evaluation

Evaluations at the dissemination phase of an intervention are characterized by a focus on the generalizability of the intervention and external validity (Urban et al., 2014). Multiple contexts, or sites, and implementations should be tested. The aim of generalizability evaluations is to ensure that the intervention performs similarly across settings, populations, or cultures depending on the goals of the intervention.

In order to test the intervention's generalizability across different settings, the final evaluation used adolescents in multiple high schools and young adults at several universities in the western United States to examine generalizability at multiple sites in a real-world format. We expected that adolescent high school students and young adult college students in the treatment group who completed four days of the fostering purpose activities would increase in searching for purpose and identified purpose. Conversely, we expected that students in the control conditions would demonstrate no increase in purpose after completing the online memory activities.

5.1. Participants

A total of 285 adolescents were recruited from multiple high schools in the western United States. Most students were randomly assigned to either the control ($n = 116$) or intervention condition ($n = 143$). However, one classroom of 27 students completed the activities during class time, meaning they could not be randomly assigned individually

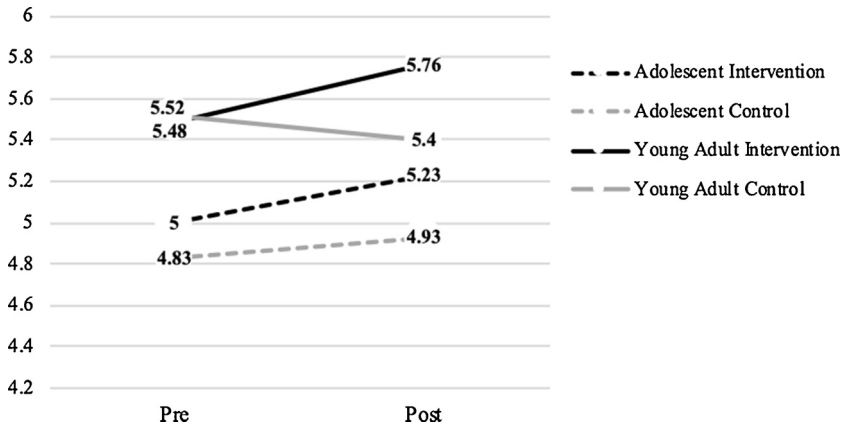

Fig. 2. SPI Group Comparisons for Adolescents and Young Adults.

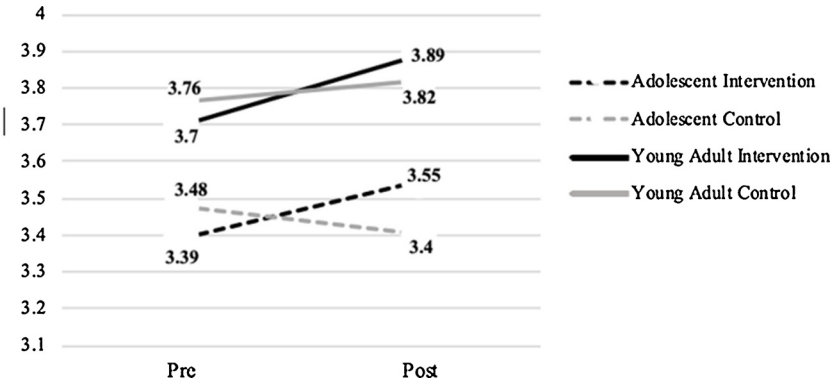

Fig. 3. CPS Group Comparisons for Adolescents and Young Adults.

and instead were all included in the intervention group, bringing the total intervention sample size to 170. The mean age for the sample was 16.83 years (SD = .93), with a range from 15 to 19. Most participants (62.2 %) identified as female, 37.1 % identified as male, and 0.6 % identified as another gender. Additionally, 29.6 % participants identified as Latino, 22.8 % as White, 17 % as Asian, 14.2 % identified as mixed race/ethnicity, and 4% Black. Students were given $5 USD for completing the pretest, fostering purpose activities, and posttest. Adolescents were included in the study if they had parental consent, or if they were 18 years of age or older and provided their own consent, and passed attention checks in the surveys.

Additionally, 294 young adults were recruited from six colleges in southern California in a separate test of the fostering purpose intervention. From the young adult sample, 121 were randomly assigned to the control group and 173 to the intervention group. The mean age for the sample was 22.12 years (SD = 3.51), with a range from 18 to 40. Most participants (83 %) identified as female. Additionally, 56.5 % of participants identified as Latino, 22.1 % as White, 8.3 % as Asian, 2.8 % as Black, and 10.3 % identified as another race or ethnicity. Students were given extra credit from their professors for participating in the study. Students were included in the study if they were 18 years of age or older, provided consent, and passed attention checks in the surveys.

5.2. Measures and procedures

To measure participants' searching for purpose we used the Searching for Purpose Inventory (Cronbach's alpha ranged from .88 to .92), which was described in study 2. To measure identified purpose in life we used the published version of the Claremont Purpose Scale (CPS; Bronk et al., 2018). The CPS consisted of 12 questions measuring each of the three components of purpose. Sample questions included: "How clear is your sense of purpose in life?" and "How important is it for you to make the world a better place in some way?" Participants responded on a five-point Likert scale. The Cronbach's alpha for this study ranged from .90 to .91.

Procedures for this study were similar to those used in study 3. Participants began by taking a pretest to measure their purpose in life and searching for purpose. After the pretest, each participant was randomly assigned to one of two conditions. Participants in the intervention and control conditions completed four sessions of activities; each session took about 30 min to complete. The intervention group completed fostering purpose activities while the control condition completed memory activities. After the four sessions, participants took the posttest survey. A lagged follow up survey was sent to participants about one week later, and again two weeks after participants completed the posttest.

5.3. Results and discussion

5.3.1. Adolescent sample

No significant differences were found in age, gender, or ethnicity between groups. Paired samples t-tests were used to test changes in purpose from the pretest to posttest in each group. Additionally, independent samples t-test were used to compare changes between the intervention group and the control group.

In paired samples t-tests, the control group showed a marginally significant decrease from the pretest (M = 3.48, SD = .59) to posttest (M = 3.40, SD = .64) on the CPS ($t(51)$ = -1.96, p = .055). There was also no significant difference in searching for purpose (SPI) from pre (M = 4.83, SD = 1.11) to posttest (M = 4.93, SD = 1.28) ($t(50)$ = 1.29, $p > .05$).

In contrast, the intervention group showed a marginally significant increase on the CPS from pretest (M = 3.39, SD = .66) to posttest (M = 3.55, SD = .84) ($t(69)$ = 1.93, p = .058), and a nonsignificant increase from pretest (M = 5.00, SD = 1.27) to posttest (M = 5.23, SD = 1.53) on the SPI ($t(68)$ = 1.62, $p > .05$).

While the change in the CPS from pre to post was marginally significant for both the control and intervention groups, when comparing groups using an independent samples t-test the change in searching for purpose (SPI) from pretest to posttest was not significantly different between the control and fostering purpose groups ($t(120)$ = 2.71, $p > .05$). However, we found a significant difference between the control and fostering purpose groups on identified purpose (CPS) ($t(120)$ = 2.39, p = .018). This shows that while the increase in purpose for the intervention group showed a small effect size, the students' change in purpose was significantly greater than it would have been without the fostering purpose intervention.

Viable validity was evaluated in the adolescent sample by examining the qualitative responses of the most important stakeholders, the teachers and students. A semi-structured interview was conducted with teachers after completing the intervention. Teachers shared positive comments about the affordability of the intervention and the ease of use with students completing activities at home, in class on mobile devices, or in computer labs. In fact, one teacher asked the authors to deliver the intervention, but after observing the ease of delivery in their first period class, this teacher volunteered to deliver the intervention to their other classes without support. Teachers also expressed that they believed the intervention was effective. They commented that on average their students articulated a clearer view of their own goals and purpose by the end of the week and thus were excited to use the intervention again in future courses. As further evidence of the intervention's viability, in the following school year more than half of the high school teachers who participated in the study used the intervention in their classes without support from the research team. Moreover, students continued to indicate positive feelings about the intervention whether they completed activities in class or on their own. While these

adolescent students did not demonstrate a large change in purpose quantitatively, their qualitative responses in activity prompts indicated that many were actively searching for purpose and many students developed a clearer purpose by the end of the intervention.

5.3.2. Young adult sample

No significant differences were found in age, gender, or ethnicity between the groups. Paired samples t-tests were performed for each measure in each group. As expected, the control group showed no significant differences from pre to post on the CPS ($t(81) = .06$, $p > .05$) or the SPI ($t(80) = -1.09$, $p > .05$).

In contrast, the intervention group showed significant increases from pre ($M = 3.75$, $SD = .59$) to posttest ($M = 3.89$, $SD = .63$) on the CPS ($t(113) = 3.21$, $p < .01$), and from pre ($M = 5.48$, $SD = 1.07$) to post ($M = 5.76$, $SD = 1.15$) on the SPI ($t(113) = 2.87$, $p < .01$).

When comparing the change in purpose between groups using an independent sample *t*-test, a significant difference was found on pre to posttest changes in the CPS ($t(194) = -2.12$, $p = .036$) and the SPI ($t(193) = -2.70$, $p = .008$). This indicated that the intervention group had significantly greater changes on both the CPS and the SPI in comparison to the control group.

These studies confirmed and replicated our earlier MTurk findings in samples of high school and college students. Specifically, we observed increases in purpose within our fostering purpose adolescents and young adult intervention groups, compared to both control groups. Furthermore, we tested a third time point after the posttest, similar to the process described in study 3. Unfortunately, too many participants failed to participate in the third follow up study and there was an insufficient sample size to run a GLM for the adolescents or young adults. Although the sample size was too small for quantitative analyses, our hypothesis appeared to be supported. The trend in mean scores on the CPS and SPI were similar to the trends found in study 3: identified purpose scores in the fostering purpose group increased and remained consistent.

These results suggest that the intervention may be more effective for young adults than for adolescents. We hypothesized two reasons for this occurrence, with the first being methodological. After initial testing with adolescents in high schools we tested many of our intervention revisions exclusively with samples of young adults through MTurk. Due to difficulties sampling minor participants using MTurk, many of the intervention iterations were tested with young adults aged 18 and older. We sampled adolescents a few times throughout the evolution of the intervention, but the methodological limitation of using MTurk so extensively may have led to the intervention evolving in a way that is more uniquely tailored to young adults rather than adolescents. The second possible reason for young adults' purpose increasing more dramatically could be contributed to young adults being more ready to identify their purpose at this point in their development. While young people can develop purpose early in adolescence, the most common time period to develop purpose is in later adolescence and early adulthood (Bronk, 2012; Pfund and Hill, 2018). Adolescents are involved in purpose exploration, but young adults tend to be identifying a purpose or committing to their purpose at a higher rate than adolescents (Bronk, Hill, Lapsley, Talib, & Finch, 2009; Bronk, 2012). The CPS measures identified purpose, and this natural course of development may be why we observed identified purpose increasing more in young adults.

The results from study four indicated that the intervention is generalizable across the contexts and sites it was tested in. In both high schools and colleges across the western United States students showed increases in searching for purpose and identified purpose as a result of the fostering purpose intervention. While this was a promising result that suggested the intervention may be generalizable to other U.S. high school and college students, it is important to remember that interventions continue to evolve. As such, we agreed with Urban et al. (2014) who recommend the further evaluation of the generalizability of an intervention across relevant contexts.

6. General discussion

The aim of this project was to build an effective purpose intervention, from the initiation to the dissemination phase. The fostering purpose project began with a few activities in the initiation phase, which were artificially selected, and modified until the project evolved into a successful, generalizable intervention that could be applied in various contexts to increase purpose in adolescents and young adults. Study one evaluated numerous activities delivered in three different combinations and demonstrated that the activities were an effective, convenient method to increase purpose in life for adolescents and young adults. Study two evaluated the standalone activities and artificially selected those most effective at increasing purpose in life, while study three tested the first full intervention which included the most effective activities and suggested that the intervention was more effective at increasing purpose when participants completed all the activities. Study four extended the context for program implementation and demonstrated the intervention's utility, practicality, and suitability in a number of high schools and college settings. This artificial selection of the program components, or activities, enabled the Fostering Purpose Project to evolve into a useful program to increase purpose in young people.

6.1. EE framework and curriculum development

Evolutionary Evaluation was not necessarily intended to guide the whole lifespan of an intervention, but may be an effective framework to use for the development of an intervention. Alignment of the evaluation and intervention phases is key. Alignment ensures that information gained through the evaluation is the most relevant for building the intervention (Urban et al., 2014). As discussed in study 1, the use of an RCT was not advised under the EE framework and our use of qualitative methods proved far more useful in that phase. Using appropriate evaluation methods in the initiation and development phases enabled the intervention to evolve and work effectively in relevant contexts. Additionally, comparing and controlling for variables helped determine when the intervention was stable. Finally, as EE recommends, testing the evaluation in multiple contexts helped us determine that the intervention could successfully be generalized to different contexts, including high school and college settings with adolescents and young adults.

One additional departure from EE was that we largely neglected viable validity in studies 2 and 3. Viable validity is important at each phase of an intervention's evolution (Urban et al., 2014) and our activities and intervention may have evolved more quickly if we had more directly evaluated viable validity at each phase. Although we measured the implementation directly in study 1 and conducted semi-structured interviews with teachers in study 4 to understand stakeholder perspectives, a greater focus and explicit measures could have revealed more about the practicality, affordability, suitability, evaluability, and helpfulness of the intervention in the contexts it was delivered in. As such, we recommend that future researchers using the EE framework to develop an intervention should focus on the full scope of viable validity throughout the evolution of their intervention (e.g., Chen, 2010).

At the conclusion of intervention's development and evaluation we discovered that our testing approach could be interpreted using the EE framework. While not tested for explicitly, the success in developing the intervention and evaluating the methods in light of EE, suggests that the EE framework could be useful in building an effective intervention from initiation to dissemination.

6.2. Conclusions

The Fostering Purpose Toolkit is a collection of four sets of free,

online activities, slightly shorter than a traditional class period. The Toolkit allows professors, teachers, parents, and adolescents flexibility in how and when they access the content or incorporate it into their curriculum. Activities can be completed inside or outside of class-time to fit the needs of students, while not overburdening teachers. Additionally, an individual could increase their purpose when completing only one or two activities, rather than the entire set, if they were limited by time or a preset curriculum. This flexibility allows for the bundling of the fostering purpose activities with other character education programs. Berkowitz and Bier (2007) found that many of the most effective character education programs combine multiple tools and program components together for more dramatic effects. Quantitative data indicated that the intervention may be most effective for young adults who are developmentally more ready for finding purpose in life (Bronk, 2012). However, qualitative and quantitative data showed that many younger adolescents were at a phase in their development where these activities brought them closer to identifying their purpose in life. We encourage teachers, professors, coaches, parents, mentors, and young people themselves to use the intervention to foster purpose. Further, we encourage others to use the EE principles to evaluate the fostering purpose intervention in other contexts, including with other cultures.

Funding

This work was supported by the John Templeton Foundation [55853, 2017].

CRediT authorship contribution statement

Brian R. Riches: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Supervision, Project administration, Funding acquisition. **Celina M. Benavides:** Conceptualization, Methodology, Software, Validation, Writing - review & editing. **Valeska X. Dubon:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - review & editing, Visualization, Funding acquisition.

Declaration of Competing Interest

The authors report no declarations of interest.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.evalprogplan.2020.101857.

References

Adolescent Moral Development Lab (2017). *Purpose in life video introduction* (May 12, 2017) Retrieved fromhttps://www.youtube.com/watch?v=5vo9LYrO8F4.

Behrend, T. S., Sharek, D. J., Meade, A. W., & Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. *Behavior Research Methods, 43*, 800–813.

Benson, P. L. (2006). *All kids are our kids: What communities must do to raise caring and responsible children and adolescents* (2nd ed.). San Francisco: Jossey-Bass.

Berkowitz, M. W., & Bier, M. C. (2007). What works in character education. *Journal of Research in Character Education, 5*, 29–49.

Bronk, K. C. (2012). A grounded theory of the development of noble youth purpose. *Journal of Adolescent Research, 27*, 78–109.

Bronk, K. C., Hill, P. L., Lapsley, D. K., Talib, T. L., & Finch, H. (2009). Purpose, hope, and life satisfaction in three age groups. *The Journal of Positive Psychology, 4*, 500–510. https://doi.org/10.1080/17439760903271439.

Bronk, K. C., Baumsteiger, R., Mangan, S., Riches, B., Dubon, V., Benavides, C. M., et al. (2019). Fostering purpose among young adults: Effective online interventions. *Journal of Character Education, 15*(2), 21–38.

Bronk, K. C., Riches, B. R., & Mangan, S. A. (2018). Claremont Purpose Scale: A measure that assesses the three dimensions of purpose among adolescents. *Research in Human Development, 15*(2), 101–117. https://doi.org/10.1080/15427609.2018.1441577.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*, 3–5.

Bundick, M. J. (2011). The benefits of reflecting on and discussing purpose in life in emerging adulthood. *New Directions in Youth Development, 132*, 89–104. https://doi.org/10.1002/yd.430.

Bundick, M. J., Yeager, D. S., King, P. E., & Damon, W. (2010). Thriving across the life span. In W. F. Overton, & R. M. Lerner (Eds.). *The handbook of life span development, vol 1: Cognition, biology, and methods* (pp. 882–923). Hoboken, NJ: John Wiley & Sons Inc. https://doi.org/10.1002/9780470880166.hlsd001024.

Chen, H. T. (2010). The bottom-up approach to integrative validity: A new perspective for program evaluation. *Evaluation and Program Planning, 33*(3), 205–214.

Damon, W. (2008). *Path to purpose: How young people find their calling in life.* New York, NY: Free Press.

Damon, W., Menon, J., & Bronk, K. C. (2003). The development of purpose during adolescence. *Applied Developmental Science, 7*, 119–128.

Dik, B. J., Steger, M. F., Gibson, A., & Peisner, W. (2011). Make your work matter: Development and pilot evaluation of a purpose-centered career education intervention. *New Directions for Youth Development, 132*, 59–73.

Frankl, V. (1984). *Man's search for meaning: Revised and updated.* New York, NY: Simon & Schuster.

Layous, K., Nelson, S. K., & Lyubomirsky, S. (2013). What is the optimal way to deliver a positive activity intervention? The case of writing about one's best possible selves. *Journal of Happiness Studies, 14*, 635–654.

Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2009). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies*Washington, DC: U.S. Department of Education, Office of Planning, Evaluation, and Policy Development. http://repository.alt.ac.uk/629/1/US_DepEdu_Final_report_2009.pdf.

Patton, M. Q. (1987). *How to use qualitative methods in evaluation (No. 4)*. Newbury Park, CA: Sage Publications, Inc.

Peterson, C., Park, N., & Seligman, M. E. P. (2005). Assessment of character strengths. In G. P. Koocher, J. C. Norcross, & S. S. HillIII (Eds.). *Psychologists' desk reference* (pp. 93–98). (2nd ed.). New York: Oxford University Press.

Pfund, G. N., & Hill, P. L. (2018). The multifaceted benefits of purpose in life. *International Forum for Logotherapy, 41*, 27–37.

Picciano, A. G., & Seaman, J. (2007). *K–12 online learning: A survey of U.S. school district administrators.* Retrieved from:Boston, MA: Sloan Consortium. https://files.eric.ed.gov/fulltext/ED530104.pdf.

Pizzolato, J. E., Brown, E. L., & Kanny, M. A. (2011). Purpose plus: Supporting youth purpose, control, and academic achievement. *New Directions for Youth Development, 132*, 75–88.

Seligman, M. E., Steen, T. A., Park, N., & Peterson, C. (2005). Positive psychology progress: Empirical validation of interventions. *The American Psychologist, 60*, 410–421.

Summerville, A., & Chartier, C. R. (2013). Pseudo-dyadic 'interaction' on Amazon's mechanical turk. *Behavior Research Methods, 45*, 116–124.

Urban, J. B., Hargraves, M., & Trochim, W. M. (2014). Evolutionary evaluation: Implications for evaluators, researchers, practitioners, funders and the evidence-based program mandate. *Evaluation and Program Planning, 45*, 127–139.

Urban, J. B., Linver, M. R., Thompson, J., Davidson, R., & Lorimer, D. (2017). Evaluating youth character development programs using evolutionary evaluation and the systems evaluation protocol. *Applied Developmental Science, 22*, 245–257.

Brian R. Riches, M.A., is a doctoral student in the Positive Developmental Psychology program at Claremont Graduate University. As an applied developmental scientist, his research interests include heroism, moral development, and purpose in life as avenues to positive youth development. He is currently engaged in creating and evaluating interventions that foster heroism and purpose in life among young people.

Dr. Celina Benavides is an Assistant Professor in the Child & Family Studies Department at California State University, Los Angeles. She earned a Ph.D. in Psychology at Claremont Graduate University, a Master's in Human Development and Psychology at Harvard, and a Master's in Education and Mathematics Credential at Whittier College. Dr. Benavides conducts research in two related topic areas: (1) the educational experiences and outcomes of students of color, and (2) the role of schools and communities in supporting the positive development of adolescents and young adults, namely through civic engagement initiatives and fostering purpose.

Valeska X. Dubon, M.A., studied Applied Positive Developmental Psychology and Evaluation at Claremont Graduate University. Her research interests focus on achievement motivation, the mentor-mentee relationship in adolescence, and the development and testing of positive interventions. She has worked on the development of mindfulness and purpose interventions, as well as the evaluation of after-school youth programs.