

# Subjective Outcome Evaluation of a Positive Youth Development Program in Mainland China

Research on Social Work Practice  
2021, Vol. 31(3) 285-297  
© The Author(s) 2020



Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/1049731520980802  
journals.sagepub.com/home/rsw



Xiaoqin Zhu<sup>1</sup>  and Daniel T. L. Shek<sup>1</sup>

## Abstract

**Objectives:** This study validated a 36-item Subjective Outcome Evaluation Scale (SOES) and examined the effectiveness of a positive youth development (PYD) program entitled “Tin Ka Ping Positive Adolescent Training through Holistic Social Programs” (TKP P.A.T.H.S. Project) and implemented in mainland China in the 2016–2017 and 2017–2018 academic years. **Methods:** We collected data from 20,480 students from 30 secondary schools in mainland China on their views toward program quality, implementer quality, and program benefits based on the SOES. **Results:** The SOES possessed good factorial, convergent, discriminant, and concurrent validities as well as internal consistency. Besides, the student respondents were generally satisfied with the program quality, implementer quality, and program benefits. Consistent with our hypothesis, junior grade students had more positive perceptions than did senior grade students. **Conclusion:** Utilizing the client satisfaction approach, this study validated the SOES and highlighted the value of PYD programs exemplified by the TKP P.A.T.H.S. Program in a non-Western context.

## Keywords

positive youth development, PYD, client satisfaction approach, adolescents, mainland China, program evaluation

Adolescent well-being is an emerging global issue deserving greater attention. According to Bruha et al. (2018), around 15% of children and adolescents showed mental health problems. In view of the rising prevalence rates of mental health problems in adolescents, one key question we should ask is how to prevent adolescent mental health issues and reduce morbidity? Traditional focus of psychiatry and clinical psychology was on adolescent weaknesses leading to “storm and stress.” However, with the emergence of positive psychology, research under the positive youth development (PYD) framework has been seeking to understand and promote the competences and potentials of young people (Shek, Dou, et al., 2019). One common assertion of different PYD models is that through building up external assets (such as family bonding and community support) and/or internal assets (such as skills and values), young people would thrive with positive developmental outcomes (Shek, Dou, et al., 2019; Shek, Lin, et al., 2020).

Review studies found that PYD programs in different formats (e.g., camp programs, arts programs, and school- and curriculum-based programs) were able to positively impact youth development in a wide range of areas such as academic achievement, self-esteem, social relationship, emotional skills, resilience, and thinking capacity (Durlak et al., 2011; Waid & Urich, 2020). For instance, in Durlak et al.’s (2011) review involving 270,034 students, it was reported that school-based PYD programs promoted students’ positive development and reduced their problem behavior. In another review based on

97,406 students who participated in socioemotional interventions in the school context, Taylor et al. (2017) highlighted that the positive outcomes of PYD programs included not only immediate educational benefits but also improved developmental trajectories over time such as graduation and safe sexual behaviors in future. Definitely, these curricula-based programs are important resources for social work in the children and youth work domain.

Nevertheless, effective PYD programs have been identified mainly in Western contexts, especially among U.S. adolescent samples. Despite a recent growing interest in PYD research in non-Western contexts such as South Africa, Brazil, and Hong Kong (Adams et al., 2019; Schwartz et al., 2017; Zhu & Shek, 2020), validated PYD programs in non-Western contexts are few. With specific reference to mainland China, Shek and Yu (2011) concluded that the number of youth prevention programs was highly inadequate after a review of PYD programs in Asia countries. This situation has not improved much after a decade. Just as Wium and Dimitrova (2019) remarked:

---

<sup>1</sup>Department of Applied Social Sciences, The Hong Kong Polytechnic University, Hong Kong, PR China

## Corresponding Author:

Daniel T. L. Shek, Department of Applied Social Sciences, The Hong Kong Polytechnic University, Hungghom, Hong Kong, PR China.  
Email: daniel.shek@polyu.edu.hk

“PYD research has mainly been conducted within the US context” (p. 2).

In addition to the lack of PYD programs and related studies in mainland China, there are three other reasons why we should develop PYD programs and conduct PYD studies in China. The first reason is the generalizability of PYD programs and research findings across different populations. In 2018, the adolescent population (aged 10–19) in mainland China was 166.86 million, roughly 13.5% of the world’s adolescent population in the same age range (UNICEF, 2019). As such, it is essential to investigate whether the existing claims, such as that PYD programs are effective and PYD research findings are valid, are applicable to Chinese adolescents as well. This point is essential as far as cultural sensitivity in social work is concerned.

The second reason is the generalizability across different cultures. While Western adolescents develop in societies where there is a strong emphasis on individual rights and personal autonomy, Chinese adolescents are under traditional Chinese socialization focusing on collectivistic interests and interdependence as well as Western influence emerged in recent decades along with rapid industrialization and globalization. Researchers pointed out that the elements of parental control, social harmony, the requirement of obedience of children, and filial piety are intrinsic to Chinese parenting (Leung & Shek, 2020). As China has a unique history and culture of more than 5,000 years, it is important to ask whether Western PYD theories and programs rooted in individualistic ideologies are applicable to Chinese adolescents.

The third reason is that with rapid urbanization and industrialization in the past four decades, risk factors impairing holistic adolescent development have accumulated, and adolescent developmental problems have intensified in mainland China. For example, Jiao et al. (2016) remarked that “the psychological health problems of children and adolescents have been more and more serious recently” in China (p. 15). Yang et al. (2015) reviewed studies on suicidal behavior in Chinese young people and reported that the lifetime prevalence of suicidal attempts was 2.7%. Tang et al. (2020) reviewed 51 studies ( $N = 144,060$  adolescents) and found that roughly one fourth of the junior high school students showed depressive symptoms. Apart from emotional distress, adolescent behavioral problem is also a major concern. In their review of 45 studies, Weng et al. (2016) remarked that juvenile delinquency has become “an emerging social problem in China” (p. 28). In view of these developmental problems, there is a need to promote inner developmental assets among adolescents, such as their competences. As such, there is an urgent need to develop validated PYD programs, which are important “in improving the lives of millions of school children” (Weissberg, 2019, p. 68).

### *Tin Ka Ping Positive Adolescent Training Through Holistic Social (TKP P.A.T.H.S.) Project in Mainland China*

A notable PYD program in the Chinese context is the P.A.T.H.S. Project, which was initially implemented in Hong

Kong to promote holistic development among youths in Hong Kong. This project incorporated 15 PYD attributes identified in effective PYD programs by Catalano et al. (2004), such as “bonding,” “emotional competence,” “resilience,” “self-efficacy,” “positive identity,” and “spirituality,” in its 120 curricula-based junior secondary teaching units. Student participants took 40 units each year during the 3-year junior secondary school learning, with each unit lasting for 30 minutes of teaching. Based on different evaluation strategies such as randomized group trial, one-group pretest–posttest design, subjective outcome evaluation developed on a client satisfaction approach, and qualitative evaluation (e.g., student diary and focus groups), findings provided support for the effectiveness of the P.A.T.H.S. Project since its commencement in 2005 (Ma & Shek, 2019; Ma, Shek, & Leung, 2019; Shek, 2019; Shek & Zhu, 2020).

As the P.A.T.H.S. Project is very successful in Hong Kong, and there is an urgent call for solid PYD programs in mainland China, the P.A.T.H.S. Project was introduced to mainland China with the financial support provided by Tin Ka Ping (TKP) Foundation. From the 2011–2012 academic year to the 2013–14 academic year, the TKP P.A.T.H.S. Project was initially implemented in four secondary schools in four cities in East China, including Shanghai, Yangzhou, Changzhou, and Suzhou. As the evaluation findings were very positive (Shek, Han, et al., 2014; Shek, Yu, et al., 2014), the project was further implemented in 30 schools in China from the 2015–2016 to the 2017–2018 academic years (i.e., the full implementation), after one-year intensive training provided to over 450 school teachers as the potential program implementers in the 2014–2015 school year. The junior secondary school curricula used in Hong Kong were adapted for local use by using simplified Chinese characters, Mandarin expressions, and teaching materials (e.g., examples and cases) that are conceptually similar and better fit into social–cultural features in mainland China. Besides, the research team further developed senior secondary school curricula, and the project was extended to senior secondary school students.

To understand the program impacts in the full implementation stage, we conducted objective outcome evaluation, subjective outcome evaluation, and qualitative evaluation. As far as objective outcome evaluation is concerned, we conducted a quasi-experimental study with an experimental group ( $n = 539$ ) and a control group ( $n = 505$ ) formed on the existing classes. Based on the data collected before and after the program implementation, findings showed that students in the experimental group performed better than did the students in the control group on measures of PYD attributes, life satisfaction, depression, and delinquency, thus providing support for the benefits of the TKP P.A.T.H.S. Project in promoting positive adolescent development and reducing developmental problems (Zhu & Shek, 2020). Qualitative evaluation relative to student diaries also revealed that students had positive evaluations of the program and teachers (Shek, Zhu, et al., 2019). Most importantly, the students perceived that the program promoted their development in different domains.

### Subjective Outcome Evaluation

In addition to the above two forms of evaluation, subjective outcome evaluation using a validated client satisfaction scale serves as an effective evaluation method. Subjective outcome evaluation assesses the perceptions of service recipients on the service received, worker(s) providing the service and impacts of the services. Such an evaluation approach based on client satisfaction has a very long history (Guttek, 1978) and has been commonly used in many domains of human services including education, psychology, social work, counseling, clinical services, and nursing (Fraser & Wu, 2016; Sozer et al., 2019). For example, in the field of higher education, validated measures of course evaluation have been developed (e.g., Chen et al., 2015; Spooen et al., 2007). The client satisfaction approach has been widely used in the social work field (Fraser & Wu, 2016; Hsieh, 2012; Posmontier et al., 2019) as well as in the high school context (Kallestrup, 2018; Pavasajjanant, 2010).

There are several advantages of conducting subjective outcome evaluation using the client satisfaction approach. First, it is cost-effective and convenient to use after completing a program. Clients can complete the questionnaire within a short period, and service providers can get first-hand information about clients' experiences in a timely manner. Second, as subjective outcome evaluation is not time-consuming and data collection does not involve high technical requirements, frontline professionals, such as teachers and social workers, can easily use this method to evaluate the service they have provided. Third, through the accumulation of client satisfaction data over time, the findings can reveal the perceived quality and effectiveness of a program under consideration (Shek & Sun, 2014; Shek, Yang, et al., 2020).

Despite these advantages and the extensive adoption of subjective outcome evaluation in different fields, the use of the client satisfaction approach in program evaluation has been challenged for its nature of subjectivity. For example, participants may assess the service based on interpersonal relationships with program implementers rather than the service quality or give positive evaluations to show their cooperation and appreciation (Shek, 2014). The most critical issue is that positive subjective outcome evaluation findings do not necessarily indicate program effectiveness. Just as Weinbach (2005) commented, "the major problem of using client-satisfaction surveys as indicators of intervention effectiveness, or of quality of a service, is that satisfaction with services and successful intervention are not the same" (p. 38).

In view of these criticisms, it is essential to enhance the objectivity of subjective outcome evaluation in data collection and analysis by using validated and reliable instruments. In fact, if validated and reliable assessment tools are used, subjective outcome evaluation can yield insightful findings that are well-correlated with objective outcome evaluation findings (Shek, 2014; Sun & Richardson, 2016). However, validated tools on subjective outcome evaluation are almost nonexistent in mainland China, especially in the secondary school sector. This major issue is closely related to the fact that published scientific papers based on subjective outcome evaluation are

rare in the Chinese context although there are ample studies in Western societies. Utilizing "client satisfaction" as a search term, a search in PsycINFO showed that there were 7,271 citations on September 14, 2020. If we added Chinese OR China, the result showed that there were only 148 citations. Besides, computer search using "subjective outcome evaluation" and Chinese OR China showed 87 pieces of work, all of which were carried out in Hong Kong. The few exceptions employed the client satisfaction approach to evaluate service effectiveness on different aspects (e.g., client satisfaction with services and service providers) in mainland China (Munro & Duckett, 2016; Xu & Du, 2018); however, the psychometric properties of the instruments were unknown.

Considering the above deficiencies in the existing literature, it is necessary to promote the client satisfaction approach in program evaluation in mainland China by using validated tools. To achieve this goal, this study attempted to examine the psychometric properties of the 36-item Subjective Outcome Evaluation Scale (SOES). We also used this scale for evaluating the TKP P.A.T.H.S. Project. Based on the conceptual framework widely used in the assessment of subjective outcomes in higher education and social services (Fraser & Wu, 2016; Shek et al., 2016; Shek, Yang, et al., 2020), the SOES used in this study had three dimensions including students' perceptions on program quality such as the design of the curriculum (10 items), instructor performance such as their attitude and mastery of teaching skills (10 items), and program benefits such as the promotion of self-confidence and resilience (16 items).

Regarding the psychometric properties of the SOES, we examined the following research question:

**Research Question 1:** Does the SOES possess acceptable psychometric properties including validity (factorial, convergent, discriminant, and concurrent validities) and reliability?

For factorial validity, we expected that there would be support for the invariance of the three-factor structure of SOES (Hypothesis 1a). We also expected that there would be support for convergent validity (Hypothesis 1b) and discriminant validity (Hypothesis 1c). For concurrent validity, it was hypothesized that ratings in the SOES would be significantly associated with three external criteria, including whether the participants would recommend the program to others (Hypothesis 1d), join similar programs again in the future (Hypothesis 1e), and overall satisfaction (Hypothesis 1f). For reliability, we expected that the total scale and subscales would have adequate reliability (Hypothesis 1g).

With reference to the results of subjective outcome evaluation, we addressed the following two research questions:

**Research Question 2:** What are the perceptions of the secondary school students on the TKP P.A.T.H.S. Project?

Based on the previous findings (Shek & Law, 2014; Shek, Zhu, et al., 2019; Zhu & Shek, 2020), it was expected that

majority of the respondents would have positive evaluations (Hypothesis 2).

**Research Question 3:** Did junior secondary school students and senior secondary school students have different perceptions of program quality, instructor performance, and program benefits?

As previous findings suggest that students at lower grades tended to have more positive views toward PYD programs than did students at the senior grades (Shek & Law, 2014), we hypothesized that junior high school students would have relatively more positive perceptions of program quality, program implementer quality, and program impacts (Hypotheses 3a, 3b, and 3c).

## Method

To evaluate student participants' perceptions of the TKP P.A.T.H.S. Project in the full implementation stage from 2015–2016 to 2017–2018 academic years, students responded to the SOES upon completion of the program each year. As the data collected in the 2015–2016 year have been analyzed and reported elsewhere (Shek, Lee, & Ma, 2018; Shek, Wu, & Law, 2018), this study utilized data collected in the 2016–2017 and 2017–2018 years.

## Participants and Procedures

At the end of the second semester each year, each project school was asked to randomly select one or more classes of students who joined the project to complete a questionnaire examining their subjective evaluations of the project. Students responded to the questionnaire voluntarily in their classrooms with the presence of a class teacher who explained study purposes, voluntary participation, and principles of anonymity and confidentiality. All the invited students agreed to complete the questionnaire. As this is part of the subject (i.e., collection of feedback at the end of the course), we asked the consent from the students. Besides, the parents agreed that their children joined the activities (including the evaluation) of the project. In total, 20,480 (10,977 in the 2016–2017 year and 9,863 in the 2017–2018 year) completed questionnaires were collected by the project schools. Among these questionnaires, 14,829 were completed by students at junior grades (Grade 7–9), and 6,011 were completed by students at senior grades (Grade 10–12).

## Measures

Modeling after existing measures on program participants' subjective evaluations on program implementation (Shek et al., 2016; Shek & Ma, 2014), we developed a 36-item "SOES" to assess student participants' views toward the project. The first 10 items measure perceived program quality, such as objectives and design of the program, on a 6-point scale (1 = *strongly disagree*; 6 = *strongly agree*). Using the same 6-point rating scale, the second 10 items measure the perceived quality of

program implementers, such as teaching attitude and skills. Finally, the remaining 16 items assess perceived program benefits (e.g., promotion of resilience and social competence) on a 5-point scale (1 = *strongly disagree*; 5 = *strongly agree*). This three-dimension framework of subjective outcome evaluation has been commonly adopted in the education and social service sectors (Fraser & Wu, 2016; Shek & Ma, 2014).

In addition to the SOES, three additional items were included to assess (1) whether the participants would recommend the program to others (1 = *definitely will not*; 4 = *definitely will*), (2) whether they would participate in similar programs in future (1 = *definitely will not*; 4 = *definitely will*), and (3) their overall satisfaction with the project (1 = *very dissatisfied*; 6 = *very satisfied*). These three items were used to check the concurrent validity of the SOES for two reasons. First, domain-specific subjective evaluations (e.g., perceptions on program quality and instructor performance) are generally regarded as conceptually related yet distinct constructs from global satisfaction (e.g., overall satisfaction with the project; Spiro et al., 2009). Second, according to the current conceptual framework of subjective outcome evaluation (Shek & Ma, 2014) and previous findings (Shek & Law, 2014; Shek, Yang, et al., 2020), participants' perceptions of the three aspects included in the SOES are key determinants of their willingness to recommend the project and participate in similar projects and overall satisfaction.

## Data Analysis

The primary psychometric property under study is the factorial validity of the SOES. We first examined the skewness and kurtosis of all items in the SOES to check normality. According to Curran et al. (1996), if requirements of normality are met (i.e., absolute values of skewness and kurtosis below 2 and 7, respectively), maximum likelihood (ML) estimation could be used in confirmatory factor analyses (CFA) with limited bias. Otherwise, ML with bootstrapping technique (2,000 resampling) can be used to reduce possible bias caused by normality violation (Arifin et al., 2012; Byrne, 2016).

In the second step, we performed CFA to test the prior three-factor structure (i.e., perceived program quality, implementer quality, and benefits) using the whole sample and different subsamples (e.g., junior secondary students and senior secondary students).

In the third step, multigroup CFA were conducted to test measurement invariance across grade level (i.e., junior vs. senior) as well as subsamples grouped based on case numbers (odd vs. even). We examined several forms of factorial invariance sequentially, including (1) configural invariance (baseline model with no constraints across subsamples), (2) weak factorial invariance (factor loadings were set to be equal across subsamples), (3) strong factorial invariance (item intercepts were further constrained to be equal), (4) structural covariance invariance (factorial variance and covariance were further set to be equal), and (5) strict factorial invariance (equality of item residuals was further imposed).

**Table 1.** Summary of Goodness of Fit for CFA Models.

| Model | Description                              | $\chi^2$  | df  | CFI | NNFI | SRMR | RMSEA [90% CI] |
|-------|--|-----------|-----|-----|------|------|----------------|
| M1    | Whole sample ( $N = 20,840$ )            | 34,464.76 | 591 | .95 | .95  | .02  | .05 [.05, .05] |
| M2    | Junior secondary sample ( $N = 14,829$ ) | 27,761.79 | 591 | .95 | .94  | .02  | .06 [.06, .06] |
| M3    | Senior secondary sample ( $N = 6,011$ )  | 12,265.69 | 591 | .94 | .93  | .03  | .06 [.06, .06] |
| M4    | Odd case number ( $N = 10,420$ )         | 18,106.95 | 591 | .95 | .95  | .02  | .05 [.05, .05] |
| M5    | Even case number ( $N = 10,420$ )        | 17,580.49 | 591 | .95 | .95  | .02  | .05 [.05, .05] |

Note. CFA = confirmatory factor analysis; CFI = comparative fit index; NNFI = Bentler-Bonett nonnormed fit index; SRMR = standardized root mean square residual; RMSEA = root mean square error of approximation; CI = confidence interval.

Several indices were employed to evaluate the overall model fit, including “comparative fit index” (CFI), “Bentler-Bonett nonnormed fit index” (NNFI), “standardized root mean square residual” (SRMR), and “root mean square error of approximation” (RMSEA). A general rule is that values of CFI and NNFI  $\geq .90$  and values of SRMR and RMSEA  $\leq .08$  indicate adequate model fit (Kline, 2015). As there is bias in the  $\chi^2$  difference test when the sample size is large, we used a more practical criterion (change in CFI, or  $\Delta$ CFI, equals to or below .01) to determine measurement invariance (Cheung & Rensvold, 2002).

The second aspect of the psychometric properties concerns the convergent and discriminant validity of the SOES. Convergent validity was assessed by “average variance extracted” (AVE) with a value equals to or higher than .50 (i.e., half or more variance in the observable items can be accounted for by the respective latent factor than residuals) indicates good convergent validity (Fornell & Larcker, 1981; Hair et al., 2006). For discriminant validity, it was examined by comparing AVEs of a pair of factors with the two factors’ shared variances indexed by the value of squared correlation. If the values of AVEs of a pair of factors are higher than their shared variance, there is support for the discriminant validity of the factors (e.g., Arifin et al., 2012).

Third, for concurrent validity, we examined the associations between participants’ evaluations toward the three facets in the SOES and their ratings on the additional three items using structural equation modeling (SEM) with bootstrapping technique (2,000 resampling). In the SEM model, three dimensions of the SOES were latent independent variables indicated by respective items, and the three satisfaction ratings were observed dependent variables.

Finally, we examined the reliability of the scale and subscales. Based on the whole sample, we examined “composite reliability” (CR) for each subscale by assessing factor loadings of the items included. According to Fornell and Larcker (1981), a value of .70 or higher indicates good reliability. Besides, internal consistency in terms of Cronbach’s  $\alpha$ s and mean inter-item correlations were calculated for each factor based on the whole sample to further check subscales’ reliability.

After clarifying the psychometric properties of the SOES, profiles of participants’ responses in SOES were also analyzed through descriptive statistics. Finally, a one-way multivariate analysis of variance was conducted to test whether junior high

school and senior high school students differed in their perceptions of the program, workers, and benefits.

## Results

### Normality Tests

For items in the subscale of perceived program quality and benefits, absolute values of skewness and kurtosis values were below 2 and 7, respectively, suggesting that these items can be regarded as following a normal distribution (Curran et al., 1996). However, absolute skewness values for the 10 items in the subscale of implementer quality were between 2.11 and 2.46, and kurtosis values of three items in this subscale were between 7.09 and 7.77. Thus, items in this subscale had moderately nonnormal distribution. Given that the normality assumption was violated, ML estimation with the bootstrapping technique (2,000 resampling) was used in CFA to reduce bias (Arifin et al., 2012).

### Psychometric Properties of the SOES

As shown in Table 1, the three-factor structure model fitted the data well (CFI  $\geq .94$ ; NNFI  $\geq .93$ ; SRMR  $\leq .02$ , and RMSEA  $\leq .06$ ), no matter which sample (the whole sample or different subsamples) was used. Therefore, the three-factor structure of the SOES was well-established.

We further tested measurement invariance across the grade levels by adding equality constraints step by step in a series of nested models (see Table 2). First, Model 6 without equality constraints well fitted the data ( $\chi^2_{(1182)} = 40,027.55$ , CFI = .94, NNFI = .94, SRMR = .02, RMSEA = .04), yielding configural invariance across grade level. Second, in Model 7, factor loadings were constrained to be equal across the grade level. Results showed significant  $\chi^2$  difference ( $\Delta\chi^2_{(33)} = 332.36$ ,  $p < .001$ ) but unchanged CFI ( $\Delta$ CFI = .000) between Models 7 and 6, supporting weak factorial invariance across grade level. Third, Model 8, where equality constraints of item intercepts were further added, showed significant chi-square difference ( $\Delta\chi^2_{(36)} = 502.84$ ,  $p < .001$ ) but a change in CFI less than .01 ( $\Delta$ CFI =  $-.001$ ) in comparison to Model 7, thereby supporting strong factorial invariance across the grade level. Fourth, equality constraints on factorial variance and covariance were further added in Model 9, which resulted in significantly different  $\chi^2$  ( $\Delta\chi^2_{(6)} = 145.23$ ,  $p < .001$ ) but unchanged

**Table 2.** Summary of Goodness of Fit for Invariance Tests.

| Model  | Description                            | $\chi^2$  | df    | CFI | NNFI | SRMR | RMSEA<br>[90% CI] | Model<br>Comparison | $\Delta\chi^2$ | $\Delta df$ | $\Delta CFI$ |
|--|--|-----------|-------|-----|------|------|-------------------|---------------------|----------------|-------------|--------------|
| Grade-level invariance                           |  |           |       |     |      |      |                   |                     |                |             |              |
| M6   | Configural invariance (baseline model) | 40,027.55 | 1,182 | .94 | .94  | .02  | .04 [.04, .04]    |                     |                |             |              |
| M7   | Weak factorial invariance              | 40,359.91 | 1,215 | .94 | .94  | .02  | .04 [.04, .04]    | M6 vs. M7           | 332.36***      | 33          | .000         |
| M8   | Strong factorial invariance            | 40,862.75 | 1,251 | .94 | .94  | .02  | .04 [.04, .04]    | M7 vs. M8           | 502.84***      | 36          | -.001        |
| M9   | Structural covariance invariance       | 41,007.98 | 1,257 | .94 | .94  | .02  | .04 [.04, .04]    | M8 vs. M9           | 145.23***      | 6           | .000         |
| M10  | Strict factorial invariance            | 43,935.75 | 1,293 | .94 | .94  | .03  | .04 [.04, .04]    | M9 vs. M10          | 2,927.77***    | 36          | -.004        |
| Subgroups invariance (odd and even case numbers) |  |           |       |     |      |      |                   |                     |                |             |              |
| M11  | Configural invariance (baseline model) | 35,687.44 | 1,182 | .95 | .95  | .02  | .04 [.04, .04]    |                     |                |             |              |
| M12  | Weak factorial invariance              | 35,751.28 | 1,215 | .95 | .95  | .02  | .04 [.04, .04]    | M11 vs. M12         | 63.84***       | 33          | .000         |
| M13  | Strong factorial invariance            | 35,792.58 | 1,251 | .95 | .95  | .02  | .04 [.04, .04]    | M12 vs. M13         | 41.30          | 36          | .000         |
| M14  | Structural covariance invariance       | 35,816.67 | 1,257 | .95 | .95  | .02  | .04 [.04, .04]    | M13 vs. M14         | 24.09***       | 6           | .000         |
| M15  | Strict factorial invariance            | 35,982.83 | 1,293 | .95 | .95  | .02  | .04 [.04, .04]    | M14 vs. M15         | 166.16***      | 36          | .000         |

Note. CFA = confirmatory factor analysis; CFI = comparative fit index; NNFI = Bentler-Bonett nonnormed fit index; SRMR = standardized root mean square residual; RMSEA = root mean square error of approximation; CI = confidence interval.

\*\*\* $p < .001$ .

CFI ( $\Delta CFI = .000$ ) as compared to Model 8. Thus, structural covariance was invariant between junior and senior high school students. Finally, equality constraints were further imposed on measurement residuals in Model 10. While  $\chi^2$  change was statistically significant ( $\Delta\chi^2_{(36)} = 2,927.77, p < .001$ ), yet the relatively small change in CFI ( $\Delta CFI = -.004$ ) supported strict factorial invariance. Taken together, measurement invariance across grade levels was supported.

With identical procedures, measurement invariance was tested between two subsamples formed based on case numbers ("odd" or "even"). Results were also depicted in Table 2. All models showed adequate goodness of model fit (CFI = .95, NNFI = .95, SRMR = .02, RMSEA = .04). Most importantly, CFI remained unchanged ( $\Delta CFI = .000$ ) when equality constraints were sequentially imposed, suggesting invariant factor loadings, item intercepts, structural covariance, and item residuals between the "odd" and "even" subsamples. In summary, the findings of the invariance tests supported factorial invariance of the three-factor structure of SOES. Thus, Hypothesis 1a was supported.

Convergent validity indicators were calculated based on CFA using the whole sample. As shown in Table 3, the "average variance extracted" (AVE) calculated based on factor loadings (all  $\geq .70$ ) was .63, .71, and .67 for the three factors, respectively, indicating that over 60% of the observable items can be accounted for by the respective latent factor than residuals. In addition, there were significant correlations between perceived program quality and implementer quality ( $r = .70, p < .001$ ), between perceived program quality and benefits ( $r = .58, p < .001$ ), and between perceived implementer quality and benefits ( $r = .49, p < .001$ ). These results suggest good convergent validity of the SOES. Thus, Hypothesis 1b was supported.

Discriminant validity was checked by comparing AVEs of any pair factors to the shared variance between them as

indicated by the squared correlation coefficient. If the AVEs are larger than the shared variance, it means that the items of the latent construct explain more variance than do the items of the other construct, suggesting adequate discriminant validity (Arifin et al., 2012; Fornell & Larcker, 1981). In this study, the aforementioned correlation coefficients ( $r$ ) between each pair of factors were .70, .58, and .49. Thus, the shared variances (i.e.,  $r^2$ ) of each two factors were .47, .34, and .24. Obviously, AVEs of any two factors (i.e., .63, .71, and .67) were larger than their shared variances, suggesting acceptable discriminant validity of the SOES. Thus, Hypothesis 1c was supported.

For the concurrent validity of the scale, we performed SEM to understand how the SOES measures predicted the three criterion measures (i.e., recommends the program to others, join similar programs, and overall satisfaction). In Table 4, the model showed adequate goodness of fit with CFI and NNFI equaling to .95 and SRMR and RMSEA below .06. Results indicated significant predictive effects of the three latent variables (i.e., "perceived program quality," "perceived implementer quality," and "perceived program benefits") on "participants' willingness to suggest other students to participate in the program" ( $R^2 = .22, p < .001$ ). The standardized regression coefficients were .15, .09, and .30, respectively ( $p < .001$  in all cases). For the rating of "participants' willingness to participate in similar programs in future," the three latent factors also demonstrated significant predicting effects ( $R^2 = .21, p < .001$ ), with standardized regression coefficients as .17, .07, and .29, respectively ( $p < .001$  in all cases). Likewise, these three factors ( $\beta = .09, .10, \text{ and } .10$ , respectively) also significantly predicted "overall satisfaction" ( $R^2 = .05, p < .001$ ). These findings provided support for Hypotheses 1d–1f.

For the reliability of the SOES measures, the CR calculated based on factor loadings was .94, .96, and .97 for the three factors, respectively, all of which were higher than .70 as a rule of thumb, thereby indicating good reliability of the scale

**Table 3.** Validity and Reliability of the Subjective Outcome Evaluation Scale.

| Subscales                     | Item    | FL  | 95% CI of FL |       | Mean FL | AVE | CR  | Cronbach's $\alpha$ | Mean Interitem Correlation |
|-------------------------------|---------|-----|--------------|-------|---------|-----|-----|---------------------|----------------------------|
|                               |         |     | Lower        | Upper |         |     |     |                     |                            |
| Perceived Program Quality     | Item 1  | .79 | .78          | .80   | .79     | .63 | .94 | .94                 | .63                        |
|                               | Item 2  | .83 | .82          | .83   |         |     |     |                     |                            |
|                               | Item 3  | .81 | .80          | .82   |         |     |     |                     |                            |
|                               | Item 4  | .78 | .77          | .78   |         |     |     |                     |                            |
|                               | Item 5  | .75 | .74          | .76   |         |     |     |                     |                            |
|                               | Item 6  | .74 | .73          | .75   |         |     |     |                     |                            |
|                               | Item 7  | .79 | .79          | .80   |         |     |     |                     |                            |
|                               | Item 8  | .82 | .81          | .82   |         |     |     |                     |                            |
|                               | Item 9  | .84 | .83          | .85   |         |     |     |                     |                            |
|                               | Item 10 | .80 | .79          | .81   |         |     |     |                     |                            |
| Perceived Implementer Quality | Item 1  | .83 | .82          | .84   | .84     | .71 | .96 | .96                 | .71                        |
|                               | Item 2  | .85 | .84          | .86   |         |     |     |                     |                            |
|                               | Item 3  | .86 | .85          | .87   |         |     |     |                     |                            |
|                               | Item 4  | .87 | .86          | .87   |         |     |     |                     |                            |
|                               | Item 5  | .86 | .85          | .86   |         |     |     |                     |                            |
|                               | Item 6  | .83 | .82          | .84   |         |     |     |                     |                            |
|                               | Item 7  | .84 | .84          | .85   |         |     |     |                     |                            |
|                               | Item 8  | .84 | .83          | .85   |         |     |     |                     |                            |
|                               | Item 9  | .82 | .81          | .83   |         |     |     |                     |                            |
|                               | Item 10 | .81 | .80          | .82   |         |     |     |                     |                            |
| Perceived Program Benefits    | Item 1  | .82 | .81          | .82   | .82     | .67 | .97 | .97                 | .67                        |
|                               | Item 2  | .84 | .84          | .85   |         |     |     |                     |                            |
|                               | Item 3  | .83 | .82          | .84   |         |     |     |                     |                            |
|                               | Item 4  | .83 | .82          | .83   |         |     |     |                     |                            |
|                               | Item 5  | .83 | .82          | .83   |         |     |     |                     |                            |
|                               | Item 6  | .82 | .81          | .82   |         |     |     |                     |                            |
|                               | Item 7  | .83 | .82          | .83   |         |     |     |                     |                            |
|                               | Item 8  | .83 | .82          | .84   |         |     |     |                     |                            |
|                               | Item 9  | .79 | .78          | .80   |         |     |     |                     |                            |
|                               | Item 10 | .83 | .82          | .83   |         |     |     |                     |                            |
|                               | Item 11 | .82 | .81          | .82   |         |     |     |                     |                            |
|                               | Item 12 | .83 | .82          | .84   |         |     |     |                     |                            |
|                               | Item 13 | .81 | .81          | .82   |         |     |     |                     |                            |
|                               | Item 14 | .81 | .80          | .82   |         |     |     |                     |                            |
|                               | Item 15 | .82 | .81          | .83   |         |     |     |                     |                            |
|                               | Item 16 | .79 | .78          | .79   |         |     |     |                     |                            |

Note. FL = factor loading; CI = confidence interval; AVE = average variance extracted; CR = composite reliability.

(Fornell & Larcker, 1981). Finally, good internal consistency was indicated by Cronbach's  $\alpha$ s ( $\geq .94$ ) and mean interitem correlations (.63–.71; see Table 3). Thus, Hypothesis 1g was supported.

**Profiles of Participants' Subjective Evaluation**

Participants' positive responses toward the three aspects in the SOES and their satisfaction ratings by grade level are summarized in Table 5–7. Over 90% of the participating students evaluated the project content positively. For instance, more than 95% of them perceived that the curriculum design is very outstanding, and they like the curriculum very much. Similarly, a very high percentage of respondents held positive perceptions of the program implementers. For instance, more than 96% of them gave positive evaluations of all the different aspects of

implementer quality (e.g., preparation, teaching skills, attitudes, and involvement). As for the perceived benefits, there were also predominant positive feedbacks. Over 90% of respondents agreed that they had benefited from program participation for all the listed aspects (e.g., resilience, self-confidence, and compassion). Student participants (over 93%) also showed very favorable ratings on the three items related to their overall satisfaction with the program. These positive evaluations supported Hypothesis 2.

**Differences in Subjective Ratings Across Grade Level**

As shown in Table 8, subjective perceptions were significantly different between students at junior and senior grades (Wilk's  $\lambda = 32.81, p < .001, \eta_p^2 = .009$ ). Specifically, students at junior grades ( $M = 5.36, SD = .80$ ) demonstrated more positive

**Table 4.** Predicting Effects of Subjective Outcome Evaluations on Satisfaction Ratings.

| Predictors  | B         | SE  | $\beta$ | 95% CI of $\beta$ |       | $R^2$ (95% CI) |
|---|-----------|-----|---------|-------------------|-------|----------------|
|   |           |     |         | Lower             | Upper |                |
| DV: Willingness to suggest other students to participate in the program |           |     |         |                   |       |                |
| Perceived program content   | .12       | .01 | .15***  | .12               | .17   | .22 (.20, .23) |
| Perceived implementer quality   | .07       | .01 | .09***  | .07               | .11   |                |
| Perceived program benefits  | .22       | .01 | .30***  | .28               | .32   |                |
| DV: Willingness to participate in similar programs in future            |           |     |         |                   |       |                |
| Perceived program content   | .14       | .01 | .17***  | .14               | .19   | .22 (.20, .23) |
| Perceived implementer quality   | .06       | .01 | .07***  | .05               | .09   |                |
| Perceived program benefits  | .22       | .01 | .29***  | .27               | .31   |                |
| DV: Overall satisfaction  |           |     |         |                   |       |                |
| Perceived program content   | .11       | .02 | .08**   | .06               | .11   | .05 (.05, .06) |
| Perceived implementer quality   | .12       | .01 | .10***  | .08               | .12   |                |
| Perceived program benefits  | .11       | .02 | .10***  | .08               | .11   |                |
|   | $\chi^2$  | df  | CFI     | NNFI              | SRMR  | RMSEA (90% CI) |
| Goodness of model fit   | 34,990.64 | 690 | .95     | .95               | .02   | .05 (.05, .05) |

Note. DV = Dependent variable, CI = Confidence interval.

\*\*\* $p < .001$ .

**Table 5.** Positive Responses (Options 4–6) Regarding Participants' Evaluations on the Program Content and Implementer Quality.

| Items   | Junior Secondary<br>(N = 14,829) |       | Senior Secondary<br>(N = 6,011) |       | Overall<br>(N = 20,840) |       |
|---|----------------------------------|-------|---------------------------------|-------|-------------------------|-------|
|   | n                                | %     | n                               | %     | n                       | %     |
| Perceived program quality   |                                  |       |                                 |       |                         |       |
| 1. The objectives of the curriculum are very clear                                      | 14,258                           | 96.15 | 5,787                           | 96.27 | 20,045                  | 96.19 |
| 2. The design of the curriculum is very good  | 14,197                           | 95.74 | 5,766                           | 95.92 | 19,963                  | 95.79 |
| 3. The activities were carefully planned  | 14,079                           | 94.94 | 5,758                           | 95.79 | 19,837                  | 95.19 |
| 4. The classroom atmosphere was very pleasant   | 13,817                           | 93.18 | 5,692                           | 94.69 | 19,509                  | 93.61 |
| 5. There was much peer interaction amongst the students                                 | 13,977                           | 94.25 | 5,739                           | 95.47 | 19,716                  | 94.61 |
| 6. I participated actively during lessons (including discussions, sharing, games, etc.) | 13,910                           | 93.80 | 5,669                           | 94.31 | 19,579                  | 93.95 |
| 7. I was encouraged to do my best   | 13,901                           | 93.74 | 5,630                           | 93.66 | 19,531                  | 93.72 |
| 8. The learning experience I encountered enhanced my interest towards the lessons       | 14,061                           | 94.82 | 5,632                           | 93.69 | 19,693                  | 94.50 |
| 9. Overall speaking, I have a very positive evaluation of the program                   | 13,937                           | 93.98 | 5,676                           | 94.43 | 19,613                  | 94.11 |
| 10. On the whole, I like this curriculum very much                                      | 14,102                           | 95.10 | 5,999                           | 99.80 | 20,101                  | 96.45 |
| Perceived quality of implementers   |                                  |       |                                 |       |                         |       |
| 1. The instructor(s) had a good mastery of the curriculum                               | 14,289                           | 96.36 | 5,803                           | 96.54 | 20,092                  | 96.41 |
| 2. The instructor(s) was well prepared for the lessons                                  | 14,367                           | 96.88 | 5,828                           | 96.96 | 20,195                  | 96.90 |
| 3. The instructor(s)' teaching skills were good   | 14,372                           | 96.92 | 5,808                           | 96.62 | 20,180                  | 96.83 |
| 4. The instructor(s) showed good professional attitudes                                 | 14,336                           | 96.68 | 5,826                           | 96.92 | 20,162                  | 96.75 |
| 5. The instructor(s) was very involved  | 14,382                           | 96.99 | 5,848                           | 97.29 | 20,230                  | 97.07 |
| 6. The instructor(s) encouraged students to participate in the activities               | 14,396                           | 97.08 | 5,849                           | 97.30 | 20,245                  | 97.14 |
| 7. The instructor(s) cared for the students   | 14,340                           | 96.70 | 5,837                           | 97.11 | 20,177                  | 96.82 |
| 8. The instructor(s) was ready to offer help to students when needed                    | 14,394                           | 97.07 | 5,802                           | 96.52 | 20,196                  | 96.91 |
| 9. The instructor(s) had much interaction with the students                             | 14,292                           | 96.38 | 5,773                           | 96.04 | 20,065                  | 96.28 |
| 10. Overall speaking, I have a very positive evaluation of the instructors              | 14,369                           | 96.90 | 5,790                           | 96.32 | 20,159                  | 96.73 |

Note. All items were rated on a 6-point Likert-type scale with 1 = strongly disagree, 2 = disagree, 3 = slightly disagree, 4 = slightly agree, 5 = agree, 6 = strongly agree. Only the positive responses (Options 4–6) are shown.

perceptions on program content than did those at senior grades ( $M = 5.30$ ,  $SD = .78$ ,  $F = 25.14$ ,  $p < .001$ ,  $\eta_p^2 = .001$ ). Similarly, junior grade students ( $M = 5.51$ ,  $SD = .73$ ) rated significantly higher than did senior grade students ( $M = 5.43$ ,  $SD = .73$ ) on

implementer quality,  $F = 54.24$ ,  $p < .001$ ,  $\eta_p^2 = .003$ . For perceived benefits, junior grade students ( $M = 4.42$ ,  $SD = .75$ ) also gave significantly higher ratings than did senior grade students ( $M = 4.29$ ,  $SD = .79$ ,  $F = 125.77$ ,  $p < .001$ ,  $\eta_p^2 = .006$ ). For the

**Table 6.** Respondents With Positive Responses (Options 3–5) Regarding Participants’ Perceived Program Benefits.

| Items   | Junior Secondary<br>(N = 14,829) |       | Senior Secondary<br>(N = 6,011) |       | Overall<br>(N = 20,840) |       |
|---|----------------------------------|-------|---------------------------------|-------|-------------------------|-------|
|   | n                                | %     | n                               | %     | n                       | %     |
| 1. It has strengthened my bonding with teachers, classmates and my family     | 13,910                           | 93.80 | 5,662                           | 94.19 | 19,572                  | 93.92 |
| 2. It has strengthened my resilience in adverse conditions                    | 14,042                           | 94.69 | 5,637                           | 93.78 | 19,679                  | 94.43 |
| 3. It has enhanced my social competence                                       | 14,122                           | 95.23 | 5,652                           | 94.03 | 19,774                  | 94.88 |
| 4. It has improved my ability in handling and expressing my emotions          | 14,106                           | 95.12 | 5,688                           | 94.63 | 19,794                  | 94.98 |
| 5. It has enhanced my cognitive competence                                    | 14,140                           | 95.35 | 5,638                           | 93.79 | 19,778                  | 94.90 |
| 6. My ability to resist harmful influences has been improved                  | 14,215                           | 95.86 | 5,647                           | 93.94 | 19,862                  | 95.31 |
| 7. It has strengthened my ability to distinguish between the good and the bad | 14,266                           | 96.20 | 5,696                           | 94.76 | 19,962                  | 95.79 |
| 8. It has increased my competence in making sensible and wise choices         | 14,206                           | 95.80 | 5,642                           | 93.86 | 19,848                  | 95.24 |
| 9. It has helped me to have life reflections                                  | 14,012                           | 94.49 | 5,590                           | 93.00 | 19,602                  | 94.06 |
| 10. It has reinforced my self-confidence                                      | 14,090                           | 95.02 | 5,626                           | 93.60 | 19,716                  | 94.61 |
| 11. It has increased my self-awareness  | 14,192                           | 95.70 | 5,722                           | 95.19 | 19,914                  | 95.56 |
| 12. It has helped me to face the future with a positive attitude              | 14,212                           | 95.84 | 5,715                           | 95.08 | 19,927                  | 95.62 |
| 13. It has helped me to cultivate compassion and care about others            | 14,115                           | 95.19 | 5,659                           | 94.14 | 19,774                  | 94.88 |
| 14. It has encouraged me to care about the community                          | 13,593                           | 91.66 | 5,433                           | 90.38 | 19,026                  | 91.30 |
| 15. It has promoted my sense of responsibility in serving the society         | 14,075                           | 94.92 | 5,632                           | 93.69 | 19,707                  | 94.56 |
| 16. It has enriched my overall development                                    | 14,334                           | 96.66 | 5,745                           | 95.57 | 20,079                  | 96.35 |

Note. All items are rated on a 5-point Likert-type scale with 1 = *unhelpful*, 2 = *not very helpful*, 3 = *slightly helpful*, 4 = *helpful*, 5 = *very helpful*. Only the positive responses (Options 3–5) are shown.

**Table 7.** Summary of the Participants’ Positive Perceptions on Satisfaction Rating.

| Items  | Junior Secondary<br>(N = 14,829) |       | Senior Secondary<br>(N = 6,011) |       | Overall<br>(N = 20,840) |       |
|--|----------------------------------|-------|---------------------------------|-------|-------------------------|-------|
|  | n                                | %     | n                               | %     | n                       | %     |
| Will you suggest other students to participate in the program <sup>a</sup> | 13,940                           | 94.00 | 5,691                           | 94.68 | 19,631                  | 94.20 |
| Will you participate in similar programs in the future <sup>a</sup>        | 13,843                           | 93.35 | 5,697                           | 94.78 | 19,540                  | 93.76 |
| Are you satisfied with the program <sup>b</sup>                            | 13,991                           | 94.35 | 5,834                           | 97.06 | 19,825                  | 95.13 |

<sup>a</sup> A 4-point Likert-type scale was used (1 = *definitely will not*, 2 = *will not*, 3 = *will*, 4 = *definitely will*). Only the positive responses (Options 3–4) are shown in this table. <sup>b</sup> A 6-point Likert-type scale was used (1 = *very dissatisfied*, 2 = *moderately dissatisfied*, 3 = *slightly dissatisfied*, 4 = *slightly satisfied*, 5 = *moderately satisfied*, 6 = *very satisfied*). Only the positive responses (Options 4–6) are shown.

**Table 8.** Participants’ Subjective Perceptions by Grade Level.

| Factors   | Junior |     | Senior |     | Comparison |            |
|---|--------|-----|--------|-----|------------|------------|
|   | M      | SD  | M      | SD  | F          | $\eta_p^2$ |
| Perceived program quality   | 5.36   | .80 | 5.30   | .78 | 32.81***   | .009       |
| Perceived implementer quality                                       | 5.51   | .73 | 5.43   | .73 | 25.14***   | .001       |
| Perceived program benefits  | 4.42   | .75 | 4.29   | .79 | 54.24***   | .003       |
| Willingness to suggest other students to participate in the program | 4.42   | .75 | 4.29   | .79 | 125.77***  | .006       |
| Willingness to participate in similar programs in future            | 3.55   | .57 | 3.46   | .61 | 100.32***  | .005       |
| Overall satisfaction  | 3.56   | .58 | 3.50   | .61 | 54.23***   | .003       |
|   | 4.78   | .94 | 4.67   | .88 | 54.55***   | .003       |

\*\*\*p < .001.

three items related to participants’ satisfaction, results also showed that junior grade students were more satisfied than senior grade students, *F* ranged between 54.23 and 100.32, *ps* < .001,  $\eta_p^2$  was between .003 and .005 (i.e., low effect sizes). The findings supported Hypotheses 3a–3c.

## Discussion and Applications to Practice

Based on a review of PYD programs in Asia countries, Shek and Yu (2011) concluded that there is a severe lack of validated youth prevention and PYD programs in mainland China.

Catalano et al. (2012) similarly pointed out that evidence-based adolescent prevention programs are sparse in non-Western contexts. Hence, the TKP P.A.T.H.S. Project represents a timely response and a pioneering program aiming to promote the holistic development of adolescents in mainland China. There is a need to evaluate the effectiveness of the program because the presence of the program does not mean that it is effective. While earlier studies support the effectiveness of the program (Shek, Han, et al., 2014; Shek, Zhu, et al., 2019), the related samples are not large. Thus, this study further provides additional robust support for the effectiveness of the TKP P.A.T.H.S. Project by analyzing subjective perceptions of a large sample of student participants. With particular reference to subjective outcome evaluation, it is noteworthy that there is no validated measure in the field in mainland China. Hence, another contribution of this study is the validation of a well-articulated SOES that can be used by social work practitioners and researchers to assess client satisfaction in the future.

Regarding the validation of the SOES, findings underscore good psychometric properties of the scale. First, the findings strongly support the three-factor conceptual framework with good factorial invariance across different subsample. Second, the scale showed good convergent and discriminant validities. Third, the concurrent validity of SOES was supported by close associations between the three factors in the SOES and the three criteria. Finally, the SOES and its subscales were internally consistent. Furthermore, the large sample of students from different grades enhanced the generalizability of the findings. These observations strongly suggest that researchers and practitioners can use the SOES objectively to assess the effectiveness of PYD programs. It is important because rigorous program evaluation using validated assessment tools is a crucial element of evidence-based practice in the field of PYD, which is grossly lacking in mainland China (Shek & Yu, 2011; Weersing, 2005).

In this study, the effectiveness of the TKP P.A.T.H.S. Project was evidenced by the student participants' overwhelmingly positive perceptions on program quality, implementer quality, and benefits of the project. These findings are consistent with the previous studies, which also showed that the project was well-received by the students (Shek, Han, et al., 2014; Shek, Wu, & Law, 2018; Shek, Zhu, et al., 2019). Several factors lead to such observations. First, as the topics (such as emotional management and social competence) were relevant to adolescent developmental needs, students were interested in the program content. Second, as we emphasized the application of the knowledge and experience gained in class, students saw the practical relevance of the program in their lives. Third, as the program emphasized experiential learning, reflective learning, and active interaction among the students and teachers, the student-centered pedagogies significantly increased students' intrinsic motivation to actively participate in the project activities. Fourth, with solid training (e.g., Shek, Leung, & Wu, 2017; Shek, Zhu, & Leung, 2017), the teachers were able to utilize interactive teaching styles, such as self-disclosure and encouragement of sharing

among the students, which further promoted students' interest and engagement in the program. These conjectures are supported by students' sharing in their diaries that they enjoyed and appreciated the learning experience in this project as the project is educative and practical and teaching is creative and stimulating (Shek, Zhu, et al., 2019).

Although both junior and senior secondary school students showed very positive perceptions of program quality, implementer, and program benefits, junior grades showed significantly more positive evaluations than did senior grades. This finding is consistent with previous evaluation results for the P.A.T.H.S. Project (e.g., Shek & Law, 2014). Indeed, Western studies also suggest that the effect sizes of youth intervention programs (e.g., social and emotional learning) were negatively correlated with participants' age (Durlak et al., 2011). January et al. (2011) also noted that early interventions were more effective than interventions with older students. There are several possible explanations for the findings. First, younger students may be more likely to regard the experiential and interactive learning activities as novel and interesting, which would increase their positive evaluation. Second, with cognitive maturation, senior grades students may favor more intellectual forms of learning, such as debates and projects. It might also be more difficult for older adolescents to engage in collaborative and reflective learning. Nevertheless, studies also show that PYD intervention effects were not dependent on participants' age (Ciocanel et al., 2017). Given the inconclusive picture, future research efforts are warranted to understand how students from different grades perceive different forms of pedagogies and whether they evaluate PYD programs in different ways.

The present findings have several important implications. First, the positive evaluations suggest that compared to the West, curriculum-based PYD interventions are equally applicable to mainland Chinese adolescents (Ciocanel et al., 2017; Durlak et al., 2011). Thus, educators and youth workers such as teachers and school social workers in mainland China are recommended to design PYD curricula and build them into routine school education. This implication is meaningful as it is essential to cultivate life skills and inner strengths among adolescents (Shek, Lin, et al., 2020), and findings have suggested the contribution of PYD attributes to positive adolescent development in mainland China (e.g., Zhou, Shek, & Zhu 2020; Zhou, Shek, Zhu, & Dou, 2020).

Second, this study also reinforces the value of subjective outcome evaluation in assessing program impacts. This study primarily provides a validated assessment tool, including three-facets that fit the commonly adopted conceptual framework in the client satisfaction approach (Fraser & Wu, 2016; Shek, Yang, et al., 2020). Although the present SOES was validated in a school educational context, it is argued that the PYD program can be generalized to other youth service contexts, such as community youth service and social work field, as long as sufficient training is provided for prospective program implementers. In this case, the SOES can also be adapted (e.g., change the wording tailor-made for the school context) and

utilized in other youth service contexts such as the social work field. In fact, the P.A.T.H.S. program in Hong Kong has been used in the community contexts by social workers (Ma & Shek, 2019; Ma, Shek, & Chen, 2019; Ma, Shek, & Leung, 2019). Besides, the subjective outcome evaluation gives program participants opportunities to express their first-hand experiences and raise their voices and, at the same time, allows program designers and implementers to have evaluation data provided by different stakeholders in time (Shek, 2013). Most importantly, this evaluation method is relatively easy and convenient to apply in the routine practice of the implementers. As it is economical and easy to administer, subjective outcome evaluation can be regarded as the initial step to be followed by randomized group trials that are costly and complicated.

Third, our findings suggest that the beneficial impacts of PYD programs may be dependent on participant characteristics such as their grade level. It is possible that younger students enjoy more benefits from their participation in PYD programs. Obviously, more studies need to be carried out to further address this issue, and educators, as well as youth workers, should pay attention to targeting students' developmental needs and features in developing and implementing PYD programs. In future studies, additional demographic and contextual factors, such as student gender, family economic status, and school climate, should be considered to further investigate the differences in responses and the contexts in which the PYD program is more or less likely to be effective.

Despite the pioneering nature of the study in mainland China, there are several limitations of the study. First, although the sample of the study is large, they were based on 30 project schools. Therefore, there is a need to implement the project and collect data from more schools in different parts of China to make the findings more generalizable. Second, although the use of three measures can be regarded as adequate, it would be helpful to use more measures to assess the criterion-related validity of the SOES. In particular, it would be theoretically and practically enlightening to examine how SOES scores predict changes in objective outcomes (such as mental health) over time. The establishment of the relationship between subjective outcome evaluation findings and objective evaluation findings (e.g., pre- and posttest comparison) will further confirm that the client satisfaction approach can serve as a reliable evaluation strategy in different fields. In addition, future research could consider examining test-retest reliability of the scale. Third, the study only collected subjective outcome evaluation data from student participants. It will draw a complete picture if data from other stakeholders, such as program implementers, can also be collected. Additionally, qualitative data can be used to shed light on subjective experiences underlying the quantitative statistics such as why junior grades gave more positive ratings than did senior grades.

### Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was conducted with the approval from the ethics review from the Human Subjects Ethics Subcommittee of The Hong Kong Polytechnic University. This research and the P.A.T.H.S. Project in mainland China are financially supported by Tin Ka Ping Foundation.

### ORCID iD

Xiaoqin Zhu  <https://orcid.org/0000-0002-5527-9729>

### References

- Adams, B. G., Wiium, N., & Abubakar, A. (2019). Developmental assets and academic performance of adolescents in Ghana, Kenya, and South Africa. *Child & Youth Care Forum, 48*, 207–222. <https://doi.org/10.1007/s10566-018-9480-z>
- Arifin, W. N., Bahri Yusoff, M. S., & Naing, N. N. (2012). Confirmatory factor analysis (CFA) of USM emotional quotient inventory (USMEQ-I) among medical degree program applicants in Universiti Sains Malaysia (USM). *Education in Medicine Journal, 4*, e26–e47. <https://doi.org/10.5959/eimj.v4i2.33>
- Bruha, L., Spyridou, V., Forth, G., & Ougrin, D. (2018). Global child and adolescent mental health: challenges and advances. *London Journal of Primary Care, 10*, 108–109. <https://doi.org/10.1080/17571472.2018.1484332>
- Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (3rd ed.). Routledge.
- Catalano, R. F., Berglund, M. L., Ryan, J. A. M., Lonczak, H. S., & Hawkins, J. D. (2004). Positive youth development in the United States: Research findings on evaluations of positive youth development programs. *The Annals of the American Academy of Political and Social Science, 591*, 98–124. <https://doi.org/10.1177/0002716203260102>
- Catalano, R. F., Fagan, A. A., Gavin, L. E., Greenberg, M. T., Irwin, C. E., Ross, D. A., & Shek, D. T. (2012). Worldwide application of prevention science in adolescent health. *The Lancet, 379*, 1653–1664. [https://doi.org/10.1016/S0140-6736\(12\)60238-4](https://doi.org/10.1016/S0140-6736(12)60238-4)
- Chen, S., McCoy, J., Cooper, L., & Lambert, T. (2015). Faculty members' perception of service-learning courses in curriculum. *International Journal of Educational and Human Developments, 1*, 9–19.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233–255. [https://doi.org/10.1207/S15328007SEM0902\\_5](https://doi.org/10.1207/S15328007SEM0902_5)
- Ciocanel, O., Power, K., Eriksen, A., & Gillings, K. (2017). Effectiveness of positive youth development interventions: A meta-analysis of randomized controlled trials. *Journal of Youth and Adolescence, 46*, 483–504. <https://doi.org/10.1007/s10964-016-0555-6>
- Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1*, 16–29. <https://doi.org/10.1037/1082-989X.1.1.16>
- Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social

- and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, 82, 405–432. <https://doi.org/10.1111/j.1467-8624.2010.01564.x>
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 39–50. <https://doi.org/10.2307/3150979>
- Fraser, M. W., & Wu, S. (2016). Measures of consumer satisfaction in social welfare and behavioral health: A systematic review. *Research on Social Work Practice*, 26, 762–776. <https://doi.org/10.1177/1049731514564990>
- Gutek, B. A. (1978). Strategies for studying client satisfaction. *Journal of Social Issues*, 34, 44–56. <https://doi.org/10.1111/j.1540-4560.1978.tb00774.x>
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* (6th ed.). Pearson Prentice Hall.
- Hsieh, C. M. (2012). Incorporating perceived importance of service elements into client satisfaction measures. *Research on Social Work Practice*, 22, 93–99. <https://doi.org/10.1177/1049731511416826>
- January, A. M., Casey, R. J., & Paulson, D. (2011). A meta-analysis of classroom-wide interventions to build social skills: Do they work? *School Psychology Review*, 40, 242–256.
- Jiao, W., Liu, L., Li, R., & Zhao, N. (2016). The current situation of child and adolescent psychological problems and intervention in China. *Asian Journal of Medical Sciences*, 7, 15–17. <https://doi.org/10.3126/ajms.v7i3.13901>
- Kallestrup, M. (2018). Learning effects of negotiation simulations: evidence from different student cohorts. In P. Bursens, V. Donche, D. Gijbels, & P. Spooren (Eds.), *Simulations of decision-making as active learning tools* (Vol. 22, pp. 165–182). Springer Nature. [https://doi.org/10.1007/978-3-319-74147-5\\_13](https://doi.org/10.1007/978-3-319-74147-5_13)
- Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). The Guilford Press.
- Leung, J. T. Y., & Shek, D. T. L. (2020). Parental control and adolescent wellbeing in Chinese adolescents in Hong Kong. *Child Indicators Research*, 13, 703–727. <https://doi.org/10.1007/s12187-019-09642-6>
- Ma, C. M. S., & Shek, D. T. L. (2019). Objective outcome evaluation of a positive youth development program: The Project P.A.T.H.S. in Hong Kong. *Research on Social Work Practice*, 29, 49–60. <https://doi.org/10.1177/1049731517711246>
- Ma, C. M. S., Shek, D. T. L., & Chen, J. M. (2019). Changes in the participants in a community-based positive youth development program in Hong Kong: Objective outcome evaluation using a one-group pretest-posttest design. *Applied Research in Quality of Life*, 14, 961–979. <https://doi.org/10.1007/s11482-018-9632-1>
- Ma, C. M. S., Shek, D. T. L., & Leung, H. (2019). Evaluation of a positive youth development program in Hong Kong: A replication. *Research on Social Work Practice*, 29, 808–819. <https://doi.org/10.1177/1049731518806579>
- Munro, N., & Duckett, J. (2016). Explaining public satisfaction with health-care systems: Findings from a nationwide survey in China. *Health Expectations*, 19, 654–666. <https://doi.org/10.1111/hex.12429>
- Pavasajjanant, N. (2010). Effectiveness of international distance education in high school between Thailand and Japan. *International Journal of Information and Communication Technology Education*, 6, 11–24. <https://doi.org/10.4018/jicte.2010070102>
- Posmontier, B., Bina, R., Glasser, S., Cinamon, T., Styr, B., & Sammarco, T. (2019). Incorporating interpersonal psychotherapy for postpartum depression into social work practice in Israel. *Research on Social Work Practice*, 29, 61–68. <https://doi.org/10.1177/1049731517707057>
- Schwartz, K. D., Theron, L. C., & Scales, P. C. (2017). Seeking and finding positive youth development among Zulu youth in South African townships. *Child Development*, 88, 1079–1086. <https://doi.org/10.1111/cdev.12869>
- Shek, D. T. L. (2013). Evaluation of the Project P.A.T.H.S. using multiple evaluation strategies. In D. T. L. Shek & R. C. F. Sun (Eds.), *Development and evaluation of positive adolescent training through Holistic Social Programs (P.A.T.H.S.)* (pp. 53–67). Springer. [https://doi.org/10.1007/978-981-4451-54-3\\_16](https://doi.org/10.1007/978-981-4451-54-3_16)
- Shek, D. T. L. (2014). Is subjective outcome evaluation related to objective outcome evaluation? Insights from a longitudinal study in Hong Kong. *Journal of Pediatric and Adolescent Gynecology*, 27, S50–S56. <https://doi.org/10.1016/j.jpag.2014.02.012>
- Shek, D. T. L. (2019). Impact of the Project P.A.T.H.S. in Hong Kong and China. *Neuropsychiatry (London)*, 9, 2217–2219. <https://doi.org/10.4172/Neuropsychiatry.1000566>
- Shek, D. T. L., Dou, D., Zhu, X., & Chai, W. (2019). Positive youth development: Current perspectives. *Adolescent Health, Medicine and Therapeutics*, 10, 131–141. <https://doi.org/10.2147/AHMT.S179946>
- Shek, D. T. L., Han, X. Y., Lee, T. Y., & Yu, L. (2014). Subjective outcome evaluation of a positive youth development program in China. *International Journal on Disability and Human Development*, 13, 275–283. <https://doi.org/10.1515/ijdh-2014-0313>
- Shek, D. T. L., & Law, M. Y. M. (2014). Evaluation of the Project P.A.T.H.S. based on the client satisfaction approach: View of the students. *Journal of Pediatric and Adolescent Gynecology*, 27, S2–S9. <https://doi.org/10.1016/j.jpag.2014.02.006>
- Shek, D. T. L., Lee, T. Y., & Ma, L. K. (2018). Subjective outcome evaluation of the Tin Ka Ping P.A.T.H.S. Project in China: View of the students. *International Public Health Journal*, 10, 71–79.
- Shek, D. T. L., Leung, J. T. Y., & Wu, J. (2017). Subjective evaluation of a training program of the Tin Ka Ping P.A.T.H.S. Project in mainland China. *International Public Health Journal*, 9, 311–323.
- Shek, D. T. L., Liang, J., & Zhu, X. (2016). Subjective outcome evaluation of a service leadership subject for university students in Hong Kong. *International Journal of Child Health and Human Development*, 9, 225–232.
- Shek, D. T. L., Lin, L., Ma, C. M. S., Yu, L., Leung, J. T. Y., Wu, F. K. Y., Leung, H., & Dou, D. (2020). Perceptions of adolescents, teachers and parents of life skills education and life skills in high school students in Hong Kong. *Applied Research in Quality of Life*. <https://doi.org/10.1007/s11482-020-09848-9>
- Shek, D. T. L., & Ma, C. M. S. (2014). Validation of a subjective outcome evaluation tool for participants in a positive youth development program in Hong Kong. *Journal of Pediatric and*

- Adolescent Gynecology*, 27, S43–S49. <https://doi.org/10.1016/j.jpap.2014.02.011>
- Shek, D. T. L., & Sun, R. C. F. (2014). Positive youth development programs for adolescents with greater psychosocial needs: Subjective outcome evaluation over 3 years. *Journal of Pediatric and Adolescent Gynecology*, 27, S17–S25. <https://doi.org/10.1016/j.jpap.2014.02.008>
- Shek, D. T. L., Wu, J., & Law, M. Y. (2018). Subjective outcome evaluation of the Tin Ka Ping P.A.T.H.S. Project: Views of senior secondary school students. *International Journal of Child and Adolescent Health*, 11, 109–117.
- Shek, D. T. L., Yang, Z., Ma, C. M. S., & Chai, C. W. Y. (2020). Subjective outcome evaluation of service-learning by the service recipients: Scale development, normative profiles and predictors. *Child Indicators Research*, 1–24. <https://doi.org/10.1007/s12187-020-09765-1>
- Shek, D. T. L., & Yu, L. (2011). A review of validated youth prevention and positive youth development programs in Asia. *International Journal of Adolescent Medicine and Health*, 23, 317–324. <https://doi.org/10.1515/IJAMH.2011.028>
- Shek, D. T. L., Yu, L., Sun, R. C. F., Lee, T. Y., Han, X. Y., Li, X. X., & Zhao, X. (2014). Objective outcome evaluation of a positive youth development program in China. *International Journal on Disability and Human Development*, 13, 255–265. <https://doi.org/10.1515/ijdh-2014-0311>
- Shek, D. T. L., & Zhu, X. (2020). Promotion of thriving among Hong Kong Chinese adolescents: Evidence from eight-wave data. *Research on Social Work Practice*, 30, 870–883. <https://doi.org/10.1177/1049731520947156>
- Shek, D. T. L., Zhu, X., & Leung, J. T. Y. (2017). Subjective outcome evaluation of the Tin Ka Ping P.A.T.H.S. project training program in mainland China. *International Journal of Child and Adolescent Health*, 10, 201–211.
- Shek, D. T. L., Zhu, X., Leung, J. T. Y., Lee, T. Y., & Wu, F. K. Y. (2019). Evaluation of the Project P.A.T.H.S. in mainland China: Findings based on student diaries. *Research on Social Work Practice*, 29, 410–419. <https://doi.org/10.1177/1049731517745994>
- Sozer, E. M., Zeybekoglu, Z., & Kaya, M. (2019). Using mid-semester course evaluation as a feedback tool for improving learning and teaching in higher education. *Assessment & Evaluation in Higher Education*, 44, 1003–1016. <https://doi.org/10.1080/02602938.2018.1564810>
- Spiro, S. E., Dekel, R., & Peled, E. (2009). Dimensions and correlates of client satisfaction: An evaluation of a shelter for runaway and homeless youth. *Research on Social Work Practice*, 19, 261–270. <https://doi.org/10.1177/1049731508329395>
- Spooren, P., Mortelmans, D., & Denekens, J. (2007). Student evaluation of teaching quality in higher education: Development of an instrument based on 10 Likert-scales. *Assessment & Evaluation in Higher Education*, 32, 667–679. <https://doi.org/10.1080/02602930601117191>
- Sun, H., & Richardson, J. T. E. (2016). Students' perceptions of the academic environment and approaches to studying in British post-graduate business education. *Assessment & Evaluation in Higher Education*, 41, 384–399. <https://doi.org/10.1080/02602938.2015.1017755>
- Tang, X., Tang, S., Ren, Z., & Wong, D. F. K. (2020). Psychosocial risk factors associated with depressive symptoms among adolescents in secondary schools in mainland China: A systematic review and meta-analysis. *Journal of Affective Disorders*, 263, 155–165. <https://doi.org/10.1016/j.jad.2019.11.118>
- Taylor, R. D., Oberle, E., Durlak, J. A., & Weissberg, R. P. (2017). Promoting positive youth development through school-based social and emotional learning interventions: A meta-analysis of follow-up effects. *Child Development*, 88, 1156–1171. <https://doi.org/10.1111/cdev.12864>
- UNICEF. (2019). *The state of the world's children 2019 statistical tables*. Retrieved June 1 from <https://data.unicef.org/resources/dataset/sowc-2019-statistical-tables/>
- Waid, J., & Uhrich, M. (2020). A scoping review of the theory and practice of positive youth development. *The British Journal of Social Work*, 50, 5–24. <https://doi.org/10.1093/bjsw/bcy130>
- Weersing, V. R. (2005). Benchmarking the effectiveness of psychotherapy: Program evaluation as a component of evidence-based practice. *Journal of the American Academy of Child and Adolescent Psychiatry*, 44, 1058–1067. <https://doi.org/10.1097/01.chi.0000172682.71384.80>
- Weinbach, R. W. (2005). *Evaluating social work services and programs*. Allyn and Bacon.
- Weissberg, R. P. (2019). Promoting the social and emotional learning of millions of school children. *Perspectives on Psychological Science*, 14, 65–69. <https://doi.org/10.1177/1745691618817756>
- Weng, X., Ran, M.-S., & Chui, W. H. (2016). Juvenile delinquency in Chinese adolescents: An ecological review of the literature. *Aggression and Violent Behavior*, 31, 26–36. <https://doi.org/10.1016/j.avb.2016.06.016>
- Wiium, N., & Dimitrova, R. (2019). Positive youth development across cultures: Introduction to the special issue. *Child & Youth Care Forum*, 48, 147–153. <https://doi.org/10.1007/s10566-019-09488-7>
- Xu, F., & Du, J. T. (2018). Factors influencing users' satisfaction and loyalty to digital libraries in Chinese universities. *Computers in Human Behavior*, 83, 64–72. <https://doi.org/10.1016/j.chb.2018.01.029>
- Yang, L.-S., Zhang, Z.-H., Sun, L., Sun, Y.-H., & Ye, D.-Q. (2015). Prevalence of suicide attempts among college students in China: A meta-analysis. *PLoS One*, 10, e0116303. <https://doi.org/10.1371/journal.pone.0116303>
- Zhu, X., & Shek, D. T. L. (2020). Impact of a positive youth development program on junior high school students in mainland China: A pioneer study. *Children and Youth Services Review*, 114, 105022. <https://doi.org/10.1016/j.childyouth.2020.105022>
- Zhou, Z., Shek, D. T. L., & Zhu, X. (2020). The importance of positive youth development attributes to life satisfaction and hopelessness in mainland Chinese adolescents. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2020.553313>
- Zhou, Z., Shek, D. T., Zhu, X., & Dou, D. (2020). Positive youth development and adolescent depression: A longitudinal study based on Mainland Chinese high school students. *International Journal of Environmental Research and Public Health*, 17, 4457. <https://doi.org/10.3390/ijerph17124457>