

RESEARCH

Open Access



Subjective outcome evaluation of a gifted education program: the Project GIFT in Hong Kong

Daniel T. L. Shek^{1*}, Alan C. K. Cheung², Anna N. N. Hui³, K. H. Leung⁴ and Ruby S. H. Cheung⁵

Abstract

Background As a pioneer gifted education program in Hong Kong, the Project GIFT in Hong Kong included intervention designed to promote psychosocial competences of the students in general (Level 1 Program) and students with high abilities (Level 2 Program). While objective outcome evaluation of the program has indicated its beneficial impact on students such as enhancement in academic performance and learning autonomy, students' perception of the program such as program's quality and instructors' performance has not been examined. The present study examined primary and secondary students' subjective perception of the programs, instructors in addition to the benefits of the Project GIFT using the client satisfaction approach.

Methods A total of 3,971 Grade 3 to Grade 9 students responded to a subjective outcome evaluation measure after program completion. In this study, we examined validity and reliability of the assessment tool as well as the response profiles of the participants using structural equation modeling and Mann-Whitney tests.

Results The findings supported the factorial, convergent, discriminant, and criterion-related validities of the assessment tool in the Level 1 and Level 2 programs. Regarding perceptions of the participants, the present results showed that students had favorable perception of the quality and benefits of the programs as well as the performance of implementers. In line with our expectation, students participating in Level 2 programs held more favorable perception of the program than did students who attended Level 1 programs.

Conclusions Consistent with our previous findings illustrating the beneficial effects of the Project GIFT on students, the present findings revealed favorable perception of students on the quality of the program and the instructors. This further underscores the effectiveness of the Project GIFT for talent development.

Keywords Gifted education, Subjective outcome evaluation, Hong Kong, Students

*Correspondence:

Daniel T. L. Shek
daniel.shek@polyu.edu.hk

¹Department of Applied Social Sciences, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong SAR

²Department of Educational Administration and Policy, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR

³Department of Social and Behavioural Sciences, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong SAR

⁴Centre for University and School Partnership, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR

⁵Department of Educational Psychology, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR



© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Introduction

Gifted education has received a lot of attention of educators and policy-makers because it helps enhance human capital and facilitate economic development in our societies. As such, a wide variety of gifted education initiatives and programs such as enrichment programs and acceleration classes have been implemented in different places over the world (e.g., [1–3]). Obviously, systematic evaluation of such programs is of paramount importance because it is crucial to the development, delivery, and justifications of the programs in addition to serving as a means for accountability [4]. Nevertheless, Callahan et al. [2] pointed out that less than 50% of the respondents from 1,566 different school districts reported there was mandatory requirement for program evaluation or any proposals to assess the quality of their gifted programs. Likewise, Zhang [5] indicated that gifted education in mainland China suffered from insufficient research on evaluation.

With reference to this deficiency, the present study was conducted to evaluate a pioneer school-based gifted education program (“Project GIFT”) in Hong Kong, which aimed to promote capability and whole-person development of “gifted” and “ordinary” students via different gifted education initiatives such as enrichment and differentiation. It is composed of three levels of programs. Level 1 program aims at nurturing three core elements of giftedness (e.g., creativity) of students in regular classrooms. Level 2 program is pull-out enrichment program targeted at promoting the capability of high ability students in different areas via immersing accelerated learning contents and affective elements into curricula. Compared to Level 1 program which is concerned about students in regular classroom, Level 2 program is more specialized and intensive in which enriched and accelerated curricula are specifically designed to match talents of high ability students in different areas such as mathematics, leadership, and creativity. For instance, the program of “Little scientists” was particularly designed to enhance inquiry skills of students proficient in Science while “Reading and creative writing classes” was implemented for highly creative students. Level 3 program aims at developing giftedness of highly talented students through individualized off-site programs. Among three levels of intervention, Level 1 and Level 2 programs were the main foci in Project GIFT involving school-based programs. It is because Level 3 programs have already been conducted by external institutions in Hong Kong such as Hong Kong Academy for Gifted Education [6]. The details of Project GIFT are described in Shek et al. [7].

To assess the quality of gifted education programs, Cao et al. [8] stated that three common methods were adopted by researchers, including the use of official standards for

gifted education, comparison of a gifted program to alternative well-established gifted counterparts (e.g. [9]), and the adoption of outcome-based evaluation. Among these three evaluation methods to gifted education programs, outcome-based evaluation is commonly utilized because many places (e.g., Finland and Denmark) are still lacking legislative policy and official standards to assess the quality and effectiveness of gifted education programs.

Conceptually, outcome-based evaluation of program intervention includes objective outcome and subjective outcome evaluation. Objective outcome evaluation is assessment of change in objective outcome indicators of the participants after program intervention. As stated by Kettler [10], objective outcome evaluation is used by educators and even economists in their assessment of gifted education programs because of close alignments of outcome-related indicators with the objectives of the programs that is crucial for assessing program effectiveness and accountability.

Although the assessment of objective outcomes has been commonly adopted by researchers to evaluate gifted education programs (e.g. [11–13]), perceptions of program participants are equally important to assess gifted education programs, especially in the area of program design and improvement. As highlighted by Borland [14], the framework of gifted education program evaluation should include perceptions of different stakeholders on the programs. Subjective outcome evaluation findings are more sensitive to the discovery of important as well as non-anticipated program outcomes that are usually neglected in objective outcome evaluation of gifted education programs. For example, Marzetta et al. [15] utilized a mixed-methods approach to assess the impact of using concepts maps on gifted learners’ understanding of ecosystem interactions in Denver. Apart from collecting pre- and post-concept map scores of students, students’ perceived benefits of using concept maps in learning science were explored via semi-structured interviewing. Likewise, Altintas and Ozdemir [16] assessed the impact of differentiation on student achievement among both gifted and non-gifted students in Istanbul. Besides adopting objective tests like Mathematics Achievement Test and Multiple Intelligence Inventory, student outcomes were also rated by teachers via answering open-ended questions in the teachers’ opinion form.

Besides qualitative semi-structured and focus group interviewing, quantitative subjective outcome evaluation is another way to collect participants’ subjective perceptions of program quality and effectiveness in gifted studies via the client satisfaction approach. Subjective outcome evaluation is a method to assess the service recipients’ views on the service obtained, quality of the implementers in providing the service, and the perceived benefit of the service [17]. It is rooted in a client

satisfaction approach that targets at assessing the extent to which participants feel satisfied with program's quality and outcomes, which is commonly used in the education sector using validated measures. For example, Sak and Ayas [18] examined the effectiveness of Education Programs for Talented Students (EPTS) Curriculum Model utilizing both objective and subjective outcome evaluations. Specifically, students' perceptions on program's content, process, product, and environment, and their overall satisfaction were assessed via the student satisfaction inventory. Ihrig et al. [19] also assessed the impact of STEM Excellence and Leadership program on students and educators utilizing a mixed-methods approach. Apart from adopting focus group interviewing, subjective outcome evaluation on the participants' perceptions of the advantages of the program, effects of program implementation on their study, and their satisfaction with the program were examined. Likewise, Alhanaya [20] investigated the experiences of college students in the university-based gifted mentorship program using a mixed-methods approach. Besides semi-structured interviewing, the researcher conducted subjective outcome evaluation to investigate students' impressions of the program and perceived benefits on their research and cognitive skills.

In short, past gifted research has highlighted the significance of employing subjective outcome evaluation in gifted education programs. Özbek and Dağyar [21] advocated that examining individuals' perceptions of the programs via program evaluation scales is regarded as effective in assessing the quality and usefulness of gifted education programs. Moreover, Sak and Ayas [18] suggested that subjective outcome evaluation of gifted education programs could help program designers to improve the programs, hence enhancing the social validity of the programs.

Although subjective outcome evaluation is regarded as an important way to evaluate gifted education programs, the adoption of client satisfaction approach underlying subjective outcome evaluation has been criticized for its subjectivity [22]. Positive feedback of program participants may be associated with the tendency to give socially desirable responses or good interpersonal relationship with the program implementers instead of quality of the programs obtained [23]. In response to the subjectivity challenge, it is necessary to enhance the objectivity of subjective outcome evaluation. Obviously, it can be attained by using assessment tools with good psychometric properties. Previous research on program evaluation has revealed that the results yielded by subjective outcome evaluation using valid and reliable measures are well correlated with objective outcome measures (e.g. [24]). Therefore, to increase the credibility of subjective outcome data in evaluating gifted education programs,

development of evaluative measures with good psychometric properties is fundamental. Nonetheless, there is a dearth of psychometrically sound measures on subjective outcome evaluation of gifted education programs, particularly in different Chinese communities. As such, we examined validity and reliability of the instruments used in Level 1 and Level 2 programs of the Project GIFT in Hong Kong [7]. These include Subjective Outcome Evaluation Scale for the Level 1 program (SOES_L1) and for the Level 2 program (SOES_L2).

In the Project GIFT, using a quasi-experimental pretest-posttest design, Shek et al. [7] showed positive changes in several objective outcome indicators (including measures in creativity, academic achievement, and self-efficacy), for students participating in both Level 1 and Level 2 programs compared to their control counterparts. Compared to students participating in Level 1 program only, students participated in both Level 1 and Level 2 programs revealed additional positive outcomes such as creativity, multiple intelligences and psychological well-being. Besides, students participating in Level 2 program usually hold more favorable perceptions on their work and learning in the program.

Although the objective outcome evaluation findings of project GIFT were generally positive, there is a lack of subjective outcome evaluation data of project GIFT. Hence, students' perceived program quality and effectiveness have not been thoroughly examined. On the one hand, even though personal-social competence (one of three core elements in gifted education) was incorporated into enriched and differentiated curricula, there was no significant change in interpersonal intelligence scores of students after participating in Project GIFT [7]. On the other hand, preliminary analysis of qualitative data from student focus groups revealed positive improvement in the interpersonal skills of students after participating in the Project. Therefore, additional evidence on the experiences of the students who participated in the Project is needed to better understand the impact of the Project on students. Unlike focus group interviews, subjective outcome evaluation enables students to assess different aspects of learning processes and outcomes in a standardized format, which in turn offers insight into the improvement of the project in a more systematic way. As such, the present study aimed to investigate the validity and reliability of the 42-item of the Level 1 subjective outcome evaluation measure (SOES_L1) and the 36-item of the tool at Level 2 (SOES_L2).

SOES_L1 and SOES_L2 were developed based on conceptual frameworks targeting at collecting students' perceptions of teaching effectiveness and hence improving the quality of teaching. Consistent with Borland's [14] utilization-focused program evaluation, the information on students' perception on characteristics and

activities of the program (like program design and teaching activities), and outcome of the program (like benefits of the program) are important for program designers to eliminate uncertainties, enhance effectiveness and make decisions regarding what those programs do and affect. Apart from characteristics and outcome of the program, Spooren et al. [25] advocated the importance of obtaining students' perception of instructors' quality because student ratings of teaching effectiveness are closely associated with the teacher who teaches the course. Based on conceptual models by Borland [14] and Spooren et al. [25], SOES_L1 and SOES_L2 were conceptualized to assess students' perception of three aspects of the program, including program quality, instructor quality and program benefits. In fact, the assessment of students' perceptions of these three aspects of the program for evaluating the course and teaching effectiveness is common in the area of education (e.g. [26, 27]).

After the development of SOES_L1 and SOES_L2, we utilized these two scales to assess subjective outcomes of students participating in Level 1 and Level 2 programs of Project GIFT. Guided by the client satisfaction approach and the conceptual framework to conduct subjective outcome evaluation in education and social service sectors (e.g. [27, 28]), SOES_L1 and SOES_L2 were used to assess three aspects of program perceptions by students. These include program quality such as curriculum design and teaching activities, implementer performance such as teaching preparation and instructional strategies, and program benefits such as self-learning abilities and learning motivation. These three aspects of perception of the program were aligned with the reaction and learning stages of the assessment of Kirkpatrick's [29] four levels model of program evaluation, in which the reaction stage of assessment indicates whether the trainees enjoy the training program or learn from it while the learning stage of assessment shows whether the trainees gain benefits in terms of knowledge, attitudes and skills from the program.

In this study, we examined the following research questions:

Construct validity and reliability of the SOES_L1 and the SOES_L2

Research Question 1: What are psychometric properties of the SOES_L1 and the SOES_L2, including their validity and reliability? Based on existing conceptual frameworks and findings (e.g. [27, 28]), we tested the following hypotheses:

- For factorial validity: we hypothesized that the three-factor model of SOES_L1 and SOES_L2 exhibited good factorial validity (Hypothesis 1a and 1b, respectively).
- Convergent validity: we expected there would be significant correlations among program quality, instructor performance and program benefits [17, 27]. Besides, the factors would explain majority of variances of their corresponding indicators in the SOES_L1 (Hypothesis 1c) and in the SOES_L2 indexed by AVE (Hypothesis 1d).
- Discriminant validity: with reference to conceptual distinctiveness [30], we hypothesized that the three-factor model would fit the data better than one-factor model and two-factor models in the SOES_L1 (Hypothesis 1e) and in the SOES_L2 (Hypothesis 1f).
- Criterion-related validity: we expected that ratings in the SOES_L1 would be significantly related to participants' willingness to participate in similar programs in future (Hypothesis 1g) and overall satisfaction (Hypothesis 1h). We also hypothesized the same for SOES_L2 with reference to future program participation (Hypothesis 1i) and satisfaction level (Hypothesis 1j).
- Reliability: we expected that the scales and subscales would have good internal consistency in the SOES_L1 (Hypothesis 1k) and the SOES_L2 (Hypothesis 1L).

Perceived effectiveness of the Project GIFT

Research Question 2: What are students' perceptions regarding Project GIFT? We hypothesized that majority of the students would have favorable evaluations of Level 1 programs (Hypothesis 2a) and Level 2 programs (Hypothesis 2b).

Differences in student perceptions between Level 1 and Level 2 programs

Research Question 3: Are the perceptions of program quality, teacher quality, and program benefits, in addition to overall satisfaction different between students who participated in Level 1 programs and Level 2 programs? Because of the more intensive nature of the Level 2 program [7], we hypothesized that students joining the Level 2 programs would have relatively more favorable perceptions of program quality, instructor quality, and program impacts, as well as overall satisfaction with the program than did Level 1 students (Hypotheses 3a, 3b, 3c, and 3d, respectively).

Methods

Participants and procedures

A total of 3,971 students participating in the survey. They were recruited from twenty primary and secondary schools in Hong Kong. They were selected based on the following selection criteria: (a) schools demonstrating strong commitment and enthusiasm for gifted education and talent development; (b) schools admitting students

with different academic achievement; (c) schools with a significant number of underprivileged students; and (d) schools from different school sponsoring bodies. Among the sample, there were 3,319 and 652 primary and secondary students, respectively, who studied from Grade 3 to Grade 9. Before collecting the data,

written informed consent from parents and assent from students were obtained. After that, students were told to answer questionnaires of subjective outcome evaluation at the end of the programs in class. Study objectives in addition to the principles of confidentiality, anonymity and voluntary participation were explained to students by their class teachers.

Instruments

The SOES_L1 and the SOES_L2 were developed in the present study. At first, a pool of items was created based on the extant literature on the participants' subjective evaluations on program implementation and effectiveness (e.g. [17, 26]). Afterward, the items were selected using expert panel reviews of the item relevance, clarity, and representativeness. Some items were deleted because they were too lengthy, lack of clarity and conciseness, questionable relevance, or undesirable similarity to other items. After incorporating the comments provided by the experts, 42 items and 36 items were retained for the SOES_L1 and the SOES_L2 to assess students' perceptions of Level 1 and Level 2 programs, respectively.

The first 18 items of SOES_L1 and 12 items of SOES_L2 were used to examine students' perceptions on the program quality such as the design and activities of the programs. The next 10 items of both SOES_L1 and SOES_L2 were utilized to assess the quality of program implementer such as implementers' knowledge of the subjects and instructional strategies. The final 12 items of both SOE scales were used to investigate perceived benefits gained from the programs such as the promotion of creativity and social competence. The three-factor model of subjective outcome evaluation which includes program quality, instructor performance and program benefits has been commonly utilized to examine participants' perception of program implementation and impact in the education settings (e.g. [26, 31]). Students were required to rate the items along a 4-point scale (1 = strongly disagree; 4 = strongly agree). Apart from the SOES_L1 and the SOES_L2, two additional items were included in the survey to examine whether participants would take part in similar programs in future (1 = definitely will not; 4 = definitely will) and overall satisfaction with the programs (1 = very dissatisfied; 4 = very satisfied).

Data analysis

There was no need to impute the missing data because of low missing rate per variable (0.8%). The results of the

Little's MCAR tests revealed that the data of SOES_L1 and SOES_L2 were missing randomly. The data of this study were treated as ordinal ones because participants responded the items in a 4-Likert point scale [32]. To assess the psychometric properties of SOES_L1 and SOES_L2, we examined the factorial validity of SOES_L1 and SOES_L2 using LISREL 8.54. Confirmatory factor analysis (CFA) using diagonally weighted least squares estimation (DWLS) was conducted to examine the fit of the three-factor model of SOES_L1 and SOES_L2 individually.

In this study, DWLS was chosen instead of the maximum likelihood estimation method (ML) because DWLS utilizes a polychoric correlation matrix in parameter estimation, which is more suitable for non-continuous variables like ordinal variables. It is likely to yield less biased and more accurate estimates of factor loadings in CFA than ML when sample size is large [32, 33]. According to previous studies of structural equation modeling [34–36], a value of comparative fit index and non-normed fit index greater than 0.95 indicates a good fit. A value of root mean square error approximation less than 0.05 is regarded as a good fit while less than 0.08 is acceptable [35, 37]. A value of standardized root mean square residual less than 0.08 is viewed as adequate [35].

Second, we focused on convergent and discriminant validity of SOES_L1 and SOES_L2. Convergent validity of these two SOE measures were assessed by average variance extracted (AVE) and mean inter-factor correlation. An AVE higher than 0.50 is adequate which reflects that the construct explains a large proportion of the variance in its indicators on average [38]. Discriminant validity of SOES_L1 was assessed by comparing the model fit of original three-factor model of SOES_L1 with four competing models, which include three 2-factor models that were formed by merging any two of the three factors into ones and a 1-factor model that were formed by merging all three factors into ones (see [30]). For instance, after merging perceived program quality and program benefits into one factor, the original three-factor model became two-factor model with program quality and benefits combined as the first factor while quality of implementer as the second factor. If this 2-factor model has better fit to the data than original 3-factor model, it means that these two merging factors are best interpreted as one unified factor and hence, these merging factors are not conceptually distinct. The same method was also applied for testing discriminant validity of SOES_L2. To evaluate the competing models, Akaike's Information Criterion and expected cross-validation index were utilized. The model with smaller values of them represented a better fit to the data [39].

Third, we tested criterion-related validity of SOES_L1 and SOES_L2 by exploring the prediction from three

Table 1 Reliability, validity and inter-factor correlations of two subjective outcome evaluation scales

Subscales	SOES_L1						SOES_L2							
	Item	Factor loading	AVE	CR	α	Mean inter-item correlation	Range of item-total correlation	Item	Factor loading	AVE	CR	α	Mean inter-item correlation	Range of item-total correlation
Perceived Program Quality	1	0.79	0.612	0.97	0.95	0.52	0.67 – 0.75	1	0.84	0.673	0.96	0.93	0.54	0.62 – 0.82
	2	0.82						2	0.87					
	3	0.80						3	0.88					
	4	0.77						4	0.78					
	5	0.77						5	0.79					
	6	0.81						6	0.87					
	7	0.73						7	0.76					
	8	0.78						8	0.79					
	9	0.77						9	0.78					
	10	0.77						10	0.76					
	11	0.79						11	0.77					
	12	0.75						12	0.93					
Perceived Implementer Quality	1	0.85	0.734	0.97	0.95	0.62	0.75 – 0.79	1	0.79	0.683	0.96	0.93	0.54	0.61 – 0.76
	2	0.84						2	0.83					
	3	0.86						3	0.85					
	4	0.85						4	0.79					
	5	0.85						5	0.77					
	6	0.86						6	0.82					
	7	0.88						7	0.83					
	8	0.85						8	0.83					
	9	0.86						9	0.89					
	10	0.85						10	0.84					
	11	0.85						11	0.78					
	12	0.88						12	0.89					
Perceived Program Benefits	1	0.82	0.667	0.96	0.94	0.58	0.70 – 0.76	1	0.80	0.694	0.96	0.94	0.58	0.67 – 0.79
	2	0.83						2	0.80					
	3	0.78						3	0.84					
	4	0.78						4	0.76					
	5	0.83						5	0.83					
	6	0.80						6	0.91					

Table 1 (continued)

	SOES_L1	SOES_L2
	7 0.79	7 0.81
	8 0.81	8 0.84
	9 0.83	9 0.88
	10 0.84	10 0.89
	11 0.85	11 0.87
	12 0.84	12 0.75
	Inter-factor correlations	
	PPQ PIQ PPB	PPQ PIQ PPB
PPQ	-	-
PIQ	0.80	0.80
PPB	0.88	0.89

All correlations are significant at 0.05 level. SOES_L1 Subjective outcome evaluation scale for Level 1 students, SOES_L2 Subjective outcome evaluation scale for Level 2 students, PPQ Perceived Program Quality, PIQ Perceived Implementer Quality, PPB Perceived Program Benefits, AVE average variance extracted, CR composite reliability

dimensions of SOE measures to two observable dependent variables – students’ willingness to participate the similar courses in future and overall level of satisfaction, in SEM models. Finally, we assessed the reliability of SOES_L1 and SOES_L2 by using composite reliability (>0.7 regarded as acceptable) [40]. In addition, we examined Cronbach’s alphas, mean inter-item correlations and average item-total correlations of the scales to illustrate internal consistency reliability for the subscales of these two SOE measures (see [41]). A value of Cronbach’s alpha more than 0.7 reflects adequate reliability while the value of inter-item correlations in-between 0.3 and 0.7 and corrected item-total correlations greater than 0.3 indicate adequate internal consistency of the scale [42].

Once both SOES_L1 and SOES_L2 were validated, profiles of students’ subjective ratings on program quality and benefits, and quality of instructor after participating in Level 1 and 2 programs were explored. Mann-Whitney tests were performed to further examine the differences in subjective ratings on the programs between Level 1 students and Level 2 counterparts.

Results

Psychometric properties of SOES_L1 and SOES_L2

Table 1 illustrates the psychometric properties of SOES_L1 and SOES_L2. The CFA results revealed that the three-factor structure of both SOES_L1 and SOES_L2 fitted the data well (SOES_L1: $\chi^2 = 5608.35$, $df = 816$, $p < .001$, NNFI = 1.00, CFI = 1.00, RMSEA = 0.041, SRMR = 0.031; SOES_L2: $\chi^2 = 772.55$, $df = 591$, $p < .001$, NNFI and CFI = 1.00, RMSEA = 0.039, SRMR = 0.055). The range of factor loadings was 0.73-0.83, 0.84-0.88, and 0.78-0.85 for three factors respectively in the SOES_L1. For the SOES_L2, the range of factor loadings was 0.76-0.93, 0.77-0.89, and 0.75-0.91, respectively. All factor loadings were significant at 0.05 level. These findings confirmed the factorial validity of SOES_L1 and SOES_L2. As such, Hypotheses 1a and 1b were supported.

Nevertheless, it is noteworthy that the NNFI and CFI indices of the three-factor structure of both SOES_L1 and SOES_L2 exhibited a perfect fit. As such, the three-factor model may be overfitting the data or may not properly identified. Hence, future research should re-visit the theoretical underpinnings of the model and/or perform factor analysis using other independent samples.

Regarding convergent validity of SOES_L1 and SOES_L2, the average variance extracted (AVE) for perceived program quality, implementer quality and program benefits was 0.61, 0.73, 0.67 and 0.67, 0.68, 0.69, for the SOES_L1 and the SOES_L2, respectively (see Table 1). The findings showed that latent factors explained over 60% of variances of corresponding indicators, which reflected an adequate convergent validity of these two SOE measures. Convergent validity of two SOE measures

were further supported by high values of average correlation coefficients among three factors ($r=.81$ and $r=.83$ for the SOES_L1 and the SOES_L2, respectively). In sum, in addition to factorial validity, convergent validity of SOES_L1 and SOES_L2 were empirically supported. These findings confirmed hypotheses 1c and 1d.

Discriminant validity of SOES_L1 and SOES_L2 were examined by comparing four competing models with original three-factor model of two SOE measures [30]. Three of the competing models were formed by merging any two of three factors into ones and hence the model became the two-factor model. For instance, after merging perceived program quality and program benefits into one factor, the original three-factor model became two-factor model with program quality and benefits combined as the first factor while quality of implementer as the second factor. The last competing model was formed by merging all three factors into ones and hence the model became the one-factor model. Table 2 indicates the fit indices of competing SEM models. The results revealed that the value of AIC and ECVI of the three-factor model were less than those of all constrained models. This reflected that the original three-factor model which specified three distinct factors fitted better to the data than constrained models (see [39]). As such, discriminant validity of SOES_L1 and SOES_L2 were empirically supported, and hence confirmed hypotheses 1e and 1f.

Criterion-related validity of SOES_L1 and SOES_L2 were assessed by testing the predictive effects of perceived program quality, implementer quality and program benefits on participants' willingness to take part in similar programs in future and their overall satisfaction. Table 3 illustrates that the predictive SEM models fitted the data well (SOES_L1: $\chi^2=6498.21$, $df=895$, $p<.001$, NNFI=1.00, CFI=1.00, RMSEA=0.043, SRMR=0.033; SOES_L2: $\chi^2=900.52$, $df=658$, $p<.001$, NNFI=1.00,

CFI=1.00, RMSEA=0.043, SRMR=0.059). The findings of this study revealed that three factors of SOES_L1 positively contributed to students' willingness to participate in similar courses again ($\beta=0.77$, $p<.05$, $\beta=0.12$, $p<.05$, $\beta=0.15$, $p<.05$ respectively for three factors, $R^2=0.36$) and their feeling of satisfaction ($\beta=0.69$, $p<.05$, $\beta=0.24$, $p<.05$, $\beta=0.12$, $p<.05$, respectively for three factors, $R^2=0.41$). As such, the findings supported Hypotheses 1g and 1h.

However, only perceived program quality significantly predicted students' willingness to participate in similar courses again ($\beta=1.94$, $p<.05$, $\beta=-0.45$, $p>.05$, $\beta=-0.73$, $p>.05$ respectively for three factors, $R^2=0.37$) and their feeling of satisfaction ($\beta=1.97$, $p<.05$, $\beta=-0.30$, $p>.05$, $\beta=-0.90$, $p>.05$, respectively for three factors, $R^2=0.38$) in the SOES_L2 case. Hypotheses 1i and 1j were only partly supported. It is noteworthy that standard errors were large (mean standard error was 0.40 and 0.42, respectively for two criterion variables) in the SOES_L2 case. It might be attributed to small sample size of SOES_L2 ($N<200$). In fact, the number of students attending the Level 2 programs was much smaller than those of Level 1 programs. It is because the Level 2 program is a school-based pull-out program designed for students with high ability in different areas. Students with high potential and abilities were identified and selected based on the school-based talent search databases, examination results, performance and awards in competitions, as well as teacher-, parent-, and self-nominations in the present study. As such, the reliability of the estimates was questionable (see [43]). Consequently, the results of predictive effects of three factors on two criterion variables for the SOES_L2 must be interpreted with caution. Future research may recruit a large sample of Level 2 students and explore the contribution of perceived program quality, implementer quality and program benefits on the participants'

Table 2 Goodness-of-fit indices of the competing models for SOES_1 and SOES_L2

	Fit indices								
	χ^2	df	p-value	NNFI	CFI	RMSEA	SRMR	AIC	ECVI
SOES_L1									
1. Three-factor model (PPQ, PIQ, PPB)	5608.4	816	<0.001	1.00	1.00	0.041	0.031	5782.4	1.66
2. Two-factor model A (PPQ+PIQ, PPB)	16211.7	818	<0.001	0.99	0.99	0.073	0.057	16381.7	4.69
3. Two-factor model B (PPQ+PPB, PIQ)	8489.2	818	<0.001	1.00	1.00	0.052	0.040	8659.2	2.48
4. Two-factor model C (PIQ+PPB, PPQ)	17773.3	818	<0.001	0.99	0.99	0.077	0.058	17943.3	5.14
5. One-factor model (PPQ+PIQ+PPB)	21829.8	819	<0.001	0.99	0.99	0.086	0.065	21997.8	6.30
SOES_L2									
6. Three-factor model (PPQ, PIQ, PPB)	772.6	591	<0.001	1.00	1.00	0.039	0.055	922.6	4.61
7. Two-factor model D (PPQ+PIQ, PPB)	1118.2	593	<0.001	1.00	1.00	0.067	0.070	1264.2	6.32
8. Two-factor model E (PPQ+PPB, PIQ)	974.5	593	<0.001	1.00	1.00	0.057	0.060	1120.5	5.60
9. Two-factor model F (PIQ+PPB, PPQ)	1209.4	593	<0.001	1.00	1.00	0.072	0.072	1355.4	6.78
10. One-factor model (PPQ+PIQ+PPB)	1374.8	594	<0.001	0.99	0.99	0.081	0.077	1518.8	7.59

PPQ Perceived Program Quality, PIQ Perceived Implementer Quality, PPB Perceived Program Benefits, NNFI non-normed fit index, CFI comparative fit index, RMSEA root-mean-square error of approximation, SRMR standardized root-mean-square residual, AIC Akaike information criterion, ECVI expected cross-validation index

Table 3 Predictions from the three factors of SOES_L1 and SOES_L2 to two satisfaction-related variables

	β	SE	χ^2	df	NNFI	CFI	RMSEA	SRMR
SOES_L1								
DV: Willingness to participate in similar programs in future								
Perceived Program Quality	0.77*	0.053	6498.21	895	1.00	1.00	0.043	0.033
Perceived Implementer Quality	0.12*	0.029						
Perceived Program Benefits	0.15*	0.048						
DV: Overall satisfaction								
Perceived Program Quality	0.69*	0.055						
Perceived Implementer Quality	0.24*	0.031						
Perceived Program Benefits	0.12*	0.054						
SOES_L2								
DV: Willingness to participate in similar programs in future								
Perceived Program Quality	1.94*	0.43	900.52	658	1.00	1.00	0.043	0.059
Perceived Implementer Quality	-0.45	0.29						
Perceived Program Benefits	-0.73	0.47						
DV: Overall satisfaction								
Perceived Program Quality	1.97*	0.45						
Perceived Implementer Quality	-0.30	0.33						
Perceived Program Benefits	-0.90	0.49						

Note. Regression coefficients are standardized ones.

* $p < .05$

motivation to take part the similar programs and the feeling of satisfaction about the programs.

Regarding the reliability of SOES_L1 and SOES_L2, mean composite reliability of three subscales was 0.97 and 0.96 in SOES_L1 and SOES_L2, respectively (see Table 1). The findings illustrated that these two SOE measures were reliable. The reliability of two SOE measures were further supported by average Cronbach's alpha among three subscales, the range of mean inter-item correlations, and the range of item-total correlations (SOES_L1: 0.95, 0.52-0.62, 0.67-0.79, respectively; SOES_L2: 0.93, 0.54-0.58, 0.61-0.82, respectively). As such, Hypotheses 1k and 1 L were supported. In sum, factorial, convergent, discriminant, and criterion validity, in addition to reliability of SOES_L1 and SOES_L2 were confirmed. As such, SOES_L1 and SOES_L2 are valid and reliable measures to assess students' perceptions of the quality of the program and the implementer, and the benefits gained from the program.

Profiles for participants' positive evaluation

Tables 4 and 5 show students' positive responses to the three factors of SOES_L1 and SOES_L2 and their satisfaction ratings. Regarding perceived quality of Level 1 programs, except the item "The program provided an environment where I can be my "true self", over 80% of students evaluated the program contents positively. For example, greater than 86% of students agreed that the activities in the program were beneficial for their learning. Apart from program content, over 82.5% of students offered positive evaluation on instructor quality. For instance, more than 90% of students agreed that teachers

had a good mastery of the program content. As for the benefits of Level 1 programs, the evaluation was also mainly positive. Over 78% of students agreed that Level 1 programs were beneficial to them in many aspects such as analytical ability and learning motivation. A large proportion of students offered favorable feedback on the overall satisfaction about the programs. For example, over 84% of students admitted that they were satisfied with the programs on the whole.

Similarly, for students who participated in Level 2 programs, they evaluated the programs favorably. Regarding the perceived program quality, over 81% of students offered positive feedback on program contents. For instance, over 93% of students agreed that the program contents were well-designed. Moreover, over 86% of students rated instructor performance positively. For example, over 95% of students agreed that teachers mastered the program contents well. As for the benefits of Level 2 programs, over 78% of students agreed that Level 2 programs were beneficial to them in various aspects such as problem-solving ability and mastery of advanced knowledge and skills. As same as the satisfaction rating of Level 1 programs, a large proportion of students felt very satisfactory about Level 2 programs. Over 93% of students agreed that they were satisfied with the programs on the whole. In sum, the participating students gave positive feedback on both Level 1 and Level 2 programs. As such, Hypotheses 2a and 2b were supported.

Table 4 Positive responses (options 3 and 4) of Level 1 students to program content and implementer quality, and benefits

	Primary School (N=3,161)		Secondary School (N=597)		Total (N=3,758)	
	n	%	n	%	n	%
Perceived Program Quality						
1. The objectives of the program were very clear.	2,892	91.5	534	89.4	3,426	91.2
2. The design of the program was very good.	2,817	89.1	507	84.9	3,324	88.5
3. The activities in the program were very interesting.	2,670	84.5	473	79.2	3,143	83.6
4. The activities in the program were useful for learning.	2,747	86.9	501	83.9	3,248	86.4
5. This program enabled me to reflect my own learning process.	2,588	81.9	468	78.4	3,056	81.3
6. I enjoyed the program.	2,639	83.5	470	78.7	3,109	82.7
7. There were many opportunities for students to exchange ideas during class.	2,619	82.9	499	83.6	3,118	83.0
8. The atmosphere was very pleasant.	2,613	82.7	495	82.9	3,108	82.7
9. The program provided an environment where I can be my "true self".	2,404	76.1	429	71.9	2,833	75.4
10. I participated actively during lessons.	2,700	85.4	465	77.9	3,165	84.2
11. I was encouraged to do my best in the program.	2,661	84.2	491	82.2	3,152	83.9
12. There were much peer interaction amongst the students.	2,595	82.1	493	82.6	3,088	82.2
13. I think I am a member of my study group.	2,731	86.4	514	86.1	3,245	86.3
14. I enjoyed the sharing amongst the students.	2,624	83.0	478	80.1	3,102	82.5
15. Group discussion is valuable to me.	2,622	82.9	484	81.1	3,106	82.7
16. I learn a lot from the members in my group.	2,545	80.5	474	79.4	3,019	80.3
17. Interaction amongst my group members makes me feel more positive to my class.	2,600	82.3	482	80.7	3,082	82.0
18. On the whole, I liked this program very much.	2,662	84.2	480	80.4	3,142	83.6
Perceived Implementer Quality						
1. Teacher(s) had a good mastery of the program content.	2,883	91.2	541	90.6	3,424	91.1
2. Teacher(s) was well prepared for the program.	2,934	92.8	538	90.1	3,472	92.4
3. Teacher(s) explained clearly and easy for me to understand.	2,857	90.4	520	87.1	3,377	89.9
4. Teacher(s) was very involved.	2,930	92.7	528	88.4	3,458	92.0
5. Teacher(s) cared me.	2,720	86.0	480	80.4	3,200	85.2
6. Teacher(s) offered help to me when I needed.	2,851	90.2	521	87.3	3,372	89.7
7. Teacher(s) had much interaction with students.	2,813	89.0	503	84.3	3,316	88.2
8. Teacher(s) encouraged me to participate in the activities.	2,792	88.3	498	83.4	3,290	87.5
9. Teacher's responses stimulated me to learn.	2,645	83.7	476	79.7	3,121	83.0
10. Teacher(s) encouraged me to express myself.	2,731	86.4	474	79.4	3,205	85.3
11. Teacher(s) provided me with a lot of opportunities to think in the program.	2,782	88.0	519	86.9	3,301	87.8
12. Overall, I have very positive evaluation of the teacher(s).	2,847	90.1	525	87.9	3,372	89.7
Perceived Program Benefits						
1. The program strengthened my ability on problem solving.	2,551	80.7	472	79.1	3,023	80.4
2. The program strengthened my analytical ability.	2,625	83.0	456	76.4	3,081	82.0
3. The program strengthened my creativity.	2,592	82.0	417	69.8	3,009	80.1
4. The program enhanced my social competence.	2,498	79.0	436	73.0	2,934	78.1
5. The program helped me develop the ability to present information in a variety of formats.	2,568	81.2	473	79.2	3,041	80.9
6. The program helped me to develop better friendships with students.	2,596	82.1	463	77.6	3,059	81.4
7. The program taught me how to become a self-directed learner. (e.g., searching information actively or selecting topics by myself)	2,523	79.8	459	76.9	2,982	79.4
8. The program helped me develop the ability to summarize important information.	2,546	80.5	456	76.4	3,002	79.9
9. The program helped me to master new knowledge and skills in the subject.	2,703	85.5	502	84.1	3,205	85.3
10. The program increased my interest to the subject.	2,545	80.5	427	71.5	2,972	79.1
11. The program enhanced my learning motivation.	2,578	81.6	446	74.7	3,024	80.5
12. The program satisfied my academic needs.	2,641	83.5	460	77.1	3,101	82.5
Positive perceptions on satisfaction rating						
1. I will you participate in similar programs in the future.	2,294	72.6	372	62.3	2,666	70.9
2. On the whole, I am satisfied with this program.	2,691	85.1	488	81.7	3,179	84.6

All items were rated on a 4-point scale (1 = strongly disagree, 2 = disagree, 3 = agree, 4 = strongly agree). Only the positive responses (options 3 and 4) are illustrated

Table 5 Positive responses (options 3 and 4) of Level 2 students to program content and implementer quality, and benefits

	Primary school (N = 158)		Secondary school (N = 55)		Total (N = 213)	
	n	%	n	%	n	%
Perceived Program Quality						
1. The objectives of the program were very clear.	150	94.9	49	89.1	199	93.4
2. The design of the program was very good.	150	94.9	49	89.1	199	93.4
3. The activities in the program were very interesting.	146	92.4	44	80.0	190	89.2
4. The activities in the program were useful for learning.	143	90.5	43	78.2	186	87.3
5. This program enabled me to reflect my own learning process.	137	86.7	42	76.4	179	84.0
6. I enjoyed the program.	143	90.5	46	83.6	189	88.7
7. There were many opportunities for students to exchange ideas during class.	138	87.3	46	83.6	184	86.4
8. The atmosphere was very pleasant.	143	90.5	48	87.3	191	89.7
9. The program provided an environment where I can be my "true self".	132	83.5	42	76.4	174	81.7
10. I participated actively during lessons.	144	91.1	42	76.4	186	87.3
11. I was encouraged to do my best in the program.	141	89.2	44	80.0	185	86.9
12. On the whole, I liked this program very much.	145	91.8	48	87.3	193	90.6
Perceived Implementer Quality						
1. Teacher(s) had a good mastery of the program content.	155	98.1	49	89.1	204	95.8
2. Teacher(s) was well prepared for the program.	156	98.7	47	85.5	203	95.3
3. Teacher(s) explained clearly and easy for me to understand.	148	93.7	47	85.5	195	91.5
4. Teacher(s) was very involved.	152	96.2	49	89.1	201	94.4
5. Teacher(s) cared me.	144	91.1	43	78.2	187	87.8
6. Teacher(s) offered help to me when I needed.	148	93.7	49	89.1	197	92.5
7. Teacher(s) had much interaction with students.	146	92.4	46	83.6	192	90.1
8. Teacher(s) encouraged me to participate in the activities.	146	92.4	47	85.5	193	90.6
9. Teacher's responses stimulated me to learn.	138	87.3	46	83.6	184	86.4
10. Teacher(s) encouraged me to express myself.	141	89.2	46	83.6	187	87.8
11. Teacher(s) provided me with a lot of opportunities to think in the program.	137	86.7	48	87.3	185	86.9
12. Overall, I have very positive evaluation of the teacher(s).	149	94.3	48	87.3	197	92.5
Perceived Program Benefits						
1. The program strengthened my ability on problem solving.	143	90.5	49	89.1	192	90.1
2. The program strengthened my analytical ability.	142	89.9	50	90.9	192	90.1
3. The program strengthened my creativity.	137	86.7	39	70.9	176	82.6
4. The program enhanced my social competence.	129	81.6	39	70.9	168	78.9
5. The program helped me develop the ability to present information in a variety of formats.	126	79.7	41	74.5	167	78.4
6. The program helped me to develop friendships with students of similar interests and abilities.	133	84.2	40	72.7	173	81.2
7. The program taught me how to become a self-directed learner. (e.g., searching information actively or selecting topics by myself)	130	82.3	38	69.1	168	78.9
8. The program helped me develop the ability to summarize important information.	127	80.4	40	72.7	167	78.4
9. The program helped me to master the advanced knowledge and skills in the subject.	142	89.9	48	87.3	190	89.2
10. The program increased my interest to the subject.	130	82.3	39	70.9	169	79.3
11. The program enhanced my learning motivation.	129	81.6	42	76.4	171	80.3
12. The program satisfied my academic needs.	131	82.9	45	81.8	176	82.6
Positive perceptions on satisfaction rating						
1. I will you participate in similar programs in the future.	137	86.7	41	74.5	178	83.6
2. On the whole, I am satisfied with this program.	147	93.0	52	94.5	199	93.4

All items were rated on a 4-point scale (1 = strongly disagree, 2 = disagree, 3 = agree, 4 = strongly agree). Only the positive responses (options 3 and 4) are illustrated

Differences in students' subjective ratings between Level 1 and Level 2 programs

Results of Mann-Whitney tests revealed that Level 2 students rated significantly higher than Level 1 students on program quality ($U(n1 = 3750, n2 = 213) = 339,088, p < .001$), program benefits ($U(n1 = 3738, n2 = 213) = 361,116, p < .05$),

willingness to take part in similar programs in future ($U(n1 = 3685, n2 = 210) = 307,401, p < .001$), and overall satisfaction ($U(n1 = 3644, n2 = 209) = 304,469, p < .001$). The effect sizes (r) of the differences in terms of program quality, program benefits, willingness to participate in similar programs in future, and overall satisfaction between Level 1

and Level 2 students were 0.06, 0.04, 0.09, and 0.09, respectively, indicating a small effect size [44]. Nevertheless, there was no difference in the rating on the implementer performance between Level 1 and Level 2 programs ($U(n_1 = 3742, n_2 = 213) = 380,567, p = .265$). As such, only hypotheses 3a, 3c, and 3d were supported.

Discussion

In response to the severe lack of research on program evaluation of gifted education in mainland China [5] and Hong Kong [7], the present study targeted at examining the effectiveness of Project GIFT utilizing the subjective outcome evaluation method. In our previous published paper, results of objective outcome evaluation have revealed that there was positive change in cognitive and affective development of students after participating in Project GIFT [7]. Nevertheless, some important outcomes such as interpersonal intelligence, positive relations with others, and life satisfaction remained unchanged despite personal-social competence and affective elements were incorporated into enriched and differentiated curricula. As such, additional information regarding program effectiveness and improvement is needed for continuous implementation of school-based gifted education into schools.

As highlighted by Borland [14], the perception of program participants is more conducive to disclose any important, non-anticipated program outcomes that are easily overlooked in objective outcome evaluation of gifted education programs. Therefore, we conducted subjective outcome evaluation to obtain additional evidence on program quality of Project GIFT. Furthermore, despite subjective outcome evaluation has been utilized to assess the impact of gifted programs (e.g. [19, 20]), its practice is lacking in mainland China and validated measures are not existent. Therefore, this study adds to the literature through the development of psychometrically sound measures for performing subjective outcome evaluation of the gifted education program according to a client satisfaction approach.

With reference to the psychometric properties of SOES_L1 and SOES_L2, the results of CFA and correlational analyses revealed that both SOES_L1 and SOES_L2 were valid and reliable. First, factorial validity of the three-factor model of SOES_L1 and SOES_L2 were empirically supported. Second, three factors, namely perceived program quality, perceived implementer quality and perceived program benefits, were clearly identified and distinguished from each other (discriminant validity), in which each factor was best represented by their corresponding observable indicators (convergent validity). Third, three factors of SOES_L1 positively predicted participants' willingness to take part in similar programs in future and their overall satisfaction (criterion-related validity), though perceived program quality was only significant predictor in the SOES_L2 cases. Last, reliability

of the scales was well-supported by acceptable values of composite reliability, Cronbach's alpha, and mean item-total correlation.

In fact, subjective outcome evaluation based on the client satisfaction approach is a common practice for program evaluation in different disciplines such as education and social work. However, it has been utilized in gifted education studies mainly in the West. In view of the deficiency in adopting a client satisfaction approach to assess the effectiveness of gifted education programs in mainland China, the validation of three-factor structure of SOES_L1 and SOES_L2 affirms the applicability of the conceptual framework underlying a client satisfaction approach to conduct subjective outcome evaluation of Hong Kong gifted education programs. In addition, since validated tools for subjective outcome evaluation is a prerequisite for the assessment of program quality and effectiveness, the above findings underscore the potential usefulness of SOES_L1 and SOES_L2 in evaluation studies of gifted education programs, especially in mainland China where gifted program evaluation is limited in its practice [5]. As highlighted by vanTassel-Baska and Brown [45], perceptual data of different stakeholders on gifted education programs are crucial to program improvement and success. As such, development of psychometrically sound instruments to assess stakeholders' subjective perceptions of gifted programs could assist in evaluation and improvement of gifted education programs.

The findings of this study revealed favorable perceptions of participants on program quality and benefits, and quality of implementers of Project GIFT, which in turn provides additional evidence to the perceived benefits gained by students and the quality of Project GIFT. These findings align with our past study utilizing objective outcome evaluation to examine the effectiveness of Project GIFT, from which positive impact on several student outcomes like creativity, academic performance, and self-efficacy has been illustrated [7]. There are several factors possibly contributing to positive feedback from students' participants on Project GIFT. First, enriched and accelerated curricula in Level 1 and Level 2 programs were well-designed which matched the talent, learning ability and interest of students. Second, professional development such as training workshops and seminars was provided to teachers so that they could implement school-based gifted education effectively. Likewise, lesson observation and after-class reflection further promoted teaching ability and competence of teachers.

Third, the implementation of gifted education was further facilitated by applying effective instructional practices such as enrichment triad activities and differentiation. These strategies were supported by previous studies which applied different gifted education programs for students (e.g. [19, 20]). Furthermore, the present findings revealed that over 78% of Level 1 and 2 students offered positive feedback on peer relationships and social

competence. These findings provide additional evidence of the students' perceived benefits of Project GIFT in enhancing personal-social competence of students after participating in the programs, which might not be reflected in students' interpersonal intelligence scores of objective outcome evaluation to Project GIFT [7]. In fact, positive perception of participating students on the quality and benefits of Project GIFT offers additional evidence to confirm the usefulness of Project GIFT in Hong Kong. This further offers research-based evidence to validate Levels 1 and 2 intervention in the three-tier implementation model of gifted education in Hong Kong [46]. Once the credibility of Project GIFT is enhanced, teachers' confidence in implementing school-based gifted education into their schools would be promoted.

Although both students who participated in Level 1 and 2 programs held positive views on program quality and benefits, and quality of implementers of Project GIFT, students who attended Level 2 programs showed significantly greater positive feedback on program quality, program benefits and level of satisfaction, in addition to a higher degree of willingness to take part in similar programs in the future. These findings are consistent with past gifted studies on the impact of pull-out differentiated Level 2 programs in gifted education on students (e.g., [11, 12, 47]). For example, Dimitriadis [48] compared the impact of pull-out gifted education programs to within-classroom provisions in primary school gifted pupils and found that pull-out programs had greater influence on students. It might be attributed to the fact that students in pull-out programs attend classes at their own pace, obtain tailor-made instruction on learning and affective activities, and encouragement and support for extension. As such, students participating in Level 2 programs showed more favorable perceptions on the intervention programs.

Nevertheless, there was no difference in students' perception of implementer performance between Level 1 and Level 2 programs. It might be attributed to the fact that effective instructional strategies for the gifted such as differentiation and enrichment were provided to teachers of both Level 1 and Level 2 programs during professional development. As stated by Chan [49], there are different characteristics and competencies of teachers of gifted learners such as capability to cater learning diversity and direct individualized learning in class. Future research should adopt different methodologies such as focus group interviews to obtain deeper understanding on students' perception of instructor performance in Level 1 and Level 2 programs.

Implications

Conceptually, favorable perception of participating students on benefits of Project GIFT provides additional evidence on the usefulness of the Level 1 and Level 2 interventions in three-tier implementation model of gifted

education in Hong Kong. Coupled with evidence-based Level 3 intervention (see [50]), three tiers of interventions in the implementation model of gifted education have been research-validated which in turn, affirms the usefulness of holistic model of service delivery to gifted students in Hong Kong. Practically, as there is a lack of evaluation studies that inform gifted education programs in Asian countries [3], positive findings of this study further reveal that Project GIFT may inform the development of similar initiatives for other schools to conduct school-based gifted education in other Asian countries. Besides, the development of validated subjective outcome evaluation measures to gifted education program reinforces the value of subjective outcome evaluation because it helps increase the objectivity of the evaluation. Although these two measures were validated in the school context, it could be used to evaluate gifted education programs conducted in non-school contexts like non-profit organizations and community centers (e.g. [11]).

Limitations

Design limitations

In this study, we solely collected cross-sectional data from subjective outcome evaluation. To increase the credibility of the present findings, future research should adopt a longitudinal design which links subjective outcomes with objective outcomes, and includes teacher and parent ratings for a more comprehensive evaluation of Project GIFT. Besides, as stated by [51], people of collectivistic cultures tend to emphasize social harmony and group well-being. They may be more other-positive face concern and less likely to provide critical feedback on program evaluation. In contrast, people of individualistic cultures are likely to prioritize personal achievements and goals, leading to more likely to express self-opinions and offer critical negative comments to the programs. As such, to reduce the impact of cultural influence on the subjective outcome evaluation of the programs, data from other sources such as teachers and parent rating should be collected apart from students' subjective evaluation. However, as the participation of the respondents were anonymous in the present study, there is no strong reason to suggest that they responded in a manner to support the program effectiveness.

Sampling limitations

The sample schools were selected since they demonstrated strong commitment and enthusiasm to gifted education. The over-representation of highly motivated schools may render a sampling bias. The present positive findings may not be able to generalize to other regular schools. Future research may recruit more schools with different degree of commitment and diverse background characteristics for investigation. Also, there were only 20 project schools in Hong Kong selected to take part in the study. Even though the number of participating students is large, future

research should conduct the program and gather information from schools in different areas of mainland China, and hence makes the present findings more generalizable. However, it is noteworthy that this is a pioneer study in China.

Measurement limitation

The criterion-related validity of SOES_L1 and SOES_L2 were examined only through their predictive effects on participating students' willingness to take the similar course in future and their overall level of satisfaction. Future research should investigate the relationships between SOES_L1 and SOES_L2, and objective outcome indicators of Project GIFT such as psychological well-being and self-efficacy (see [7]). This approach can offer more robust evidence for the construct validity of SOES_L1 and SOES_L2.

Conclusions

This study serves as a method of triangulation to objective outcome evaluation study of the Project GIFT in Hong Kong [7]. This helps to supplement the research findings based on objective outcome evaluation and hence enhances the applicability of Project GIFT to inform other Asian countries about similar initiatives in school-based gifted education [52]. The evaluation findings of the present investigation strongly suggest that SOES_L1 and SOES_L2 are valid and reliable and they could be utilized to evaluate the quality of gifted education program from students' perspectives. In addition, as this gifted education program has been well received by students, this promotes the social validity of the Project GIFT which offer benefits to gifted and talented students in Hong Kong. Of course, we have to take into account the limitations of subjective outcome evaluation when we interpret the findings.

Abbreviations

AIC	Akaike's Information Criterion
AVE	Average variance extracted
CFA	Confirmatory factor analysis
ECVI	Expected cross-validation index
SEM	Structural equation modeling
SOE	Subjective outcome evaluation
SOES_L1	Subjective Outcome Evaluation Scale for the Level 1 program
SOES_L2	Subjective Outcome Evaluation Scale for the Level 2 program

Acknowledgements

The work described in this paper is based on a project entitled "Jockey Club 'Giftedness into Flourishing Talents' Project" solely supported by the Hong Kong Jockey Club Charities Trust. We would like to express our gratitude for the contribution of Professor D. W. Chan, Dr. L. K. Chan and other researchers of the Program of the Gifted and Talented for their contribution to the development and implementation of the project up to December 2018.

Authors' contributions

DTLS and ANNH developed the conceptual model and the methodology of the study. KHL analyzed the data and wrote the draft of the manuscript. DTLS, ACKC, ANNH, KHL and RSHC reviewed and checked the manuscript. DTLS, ACKC and ANNH supervised the study and were responsible for funding acquisition. DTLS, ACKC and RSHC were responsible for project administration. All authors reviewed the manuscript.

Funding

This research was funded by the Hong Kong Jockey Club Charities Trust. The writing of the paper is financially supported by Research Matching Grants (ZH4Q and ZECL) and Li and Fung Endowed Professorship in Service Leadership Education.

Data availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request. Also, the items of the questionnaires included in the paper are translated ones. Any interested readers can contact the corresponding author to get the original Chinese questionnaires.

Declarations

Ethics approval and consent to participate

The study was approved by the Survey and Behavioural Ethics Section of the Research Ethics Committee of The Chinese University of Hong Kong on 18 August 2017 (Approval Code: 6904342). The preparation of this paper is financially supported by Li and Fung Endowed Professorship in Service Leadership Education and research grants of The Hong Kong Polytechnic University (ZZUE and W02W). Informed consent was obtained from all subjects involved in the study.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 29 August 2024 / Accepted: 17 February 2026

Published online: 23 February 2026

References

1. Reid E, Boettger H. Gifted education in various countries of Europe. *Slavonic Pedagogical Stud J*. 2015;4(2):158–71. <https://doi.org/10.18355/PG.2015.4.2.158-71>.
2. Callahan CM, Moon TR, Oh S. Describing the status of programs for the gifted: a call for action. *J Educ Gifted*. 2017;40(1):20–49. <https://doi.org/10.1177/0162353216686215>.
3. Chan DW. Gifted education in Asia. In: Pfeiffer SI, Shaunessy-Dedrick E, Foley-Nicpon M, editors. *APA Handbook of Giftedness and Talent*. Washington: American Psychological Association; 2018. pp. 71–84.
4. Hodges J, Crutchfield K, Lee LE. Equitable access, locale, funding, and the alignment to performance ratings in Texas gifted programs. *J Educ Gifted*. 2021;44(1):3–30. <https://doi.org/10.1177/0162353220978306>.
5. Zhang Z. Gifted education in China. *Cogent Educ*. 2017;4:1364881. <https://doi.org/10.1080/2331186X.2017.1364881>.
6. Chan E. Can HK education system produce excellent world-class students? *China Daily*. 2023. <https://www.chinadailyhk.com/article/347154>. Accessed 17 Jan 2024.
7. Shek DTL, Cheung ACK, Hui ANN, Leung KH, Cheung RSH. Development and evaluation of a pioneer school-based gifted education program (Project GIFT) for primary and secondary students in Hong Kong. *Int J Environ Res Public Health*. 2022;19:4832. <https://doi.org/10.3390/ijerph19084832>.
8. Cao TH, Jung JY, Lee J. Assessment in gifted education: a review of the literature from 2005 to 2016. *J Adv Acad*. 2017;28(3):163–203. <https://doi.org/10.1177/1932202X17714572>.
9. Assouline SG, Lupkowski-Shoplik A. The talent search model of gifted identification. *J Psychoeduc Assess*. 2012;30(1):45–59. <https://doi.org/10.1177/0734282911433946>.
10. Kettler T. Why are economists evaluating the impact of gifted education? *J Adv Acad*. 2016;27(2):81–9. <https://doi.org/10.1177/1932202X16636172>.
11. Golle J, Zettler I, Rose N, Trautwein U, Hasselhorn M, Nagengast B. Effectiveness of a Grass Roots statewide enrichment program for gifted elementary school children. *J Res Educ Eff*. 2018;11:375–408. <https://doi.org/10.1080/19345747.2017.1402396>.

12. Gubbels J, Segers E, Verhoeven L. Cognitive, socioemotional, and attitudinal effects of triarchic enrichment program for gifted children. *J Educ Gift.* 2014;37:378–97. <https://doi.org/10.1177/0162353214552565>.
13. Shaunessy-Dedrick E, Evans L, Ferron J, Lindo M. Effects of differentiated reading on elementary students' reading comprehension and attitudes toward reading. *Gift Child Q.* 2015;59:91–107. <https://doi.org/10.1177/0016986214568718>.
14. Borland JH. Evaluating gifted programs. In: Colangelo N, Davis G, editors. *Handbook of gifted education.* 2nd ed. Boston: Allyn & Bacon; 1997. pp. 253–66.
15. Marzetta K, Mason H, Wee B. Sometimes they are fun and sometimes they are not: concept mapping with English Language acquisition (ELA) and gifted/ talented (GT) elementary students learning science and sustainability. *Educ Sci.* 2018;8(1):1–12. <https://doi.org/10.3390/educsci8010013>.
16. Altintas E, Ozdemir AS. Evaluating a newly developed differentiation approach in terms of student achievement and teachers' opinions. *Educ Sci-Theor Pract.* 2015;15(4):1103–8. <https://doi.org/10.12738/estp.2015.4.2540>.
17. Zhu X, Shek DTL. Subjective outcome evaluation of a positive youth development program in mainland China. *Res Soc Work Pract.* 2021;31(3):1–13. <https://doi.org/10.1177/1049731520980802>.
18. Sak U, Ayas B. EPTS Curriculum Model: optimum curriculum differentiator for the education of gifted students. *Gifted Educ Int.* 2020;36(2):154–69. <https://doi.org/10.1177/0261429420917879>.
19. Ihrig LM, Lane E, Mahatmya D, Assouline SG. STEM Excellence and Leadership program: increasing the level of STEM challenge and engagement for high-achieving students in economically disadvantaged rural communities. *J Educ Gifted.* 2018;41(1):24–42. <https://doi.org/10.1177/0162353217745158>.
20. Alhanaya M. Reflecting on the role of mentorship in gifted education: mixed approach study. *J Educ Gifted Young Scientists.* 2020;8(3):1261–74. <https://doi.org/10.17478/jegy.785567>.
21. Özbek G, Dağyar M. Examining gifted students' evaluations of their education programs in terms of their project production and management. *Front Psychol.* 2022;13:833395. <https://doi.org/10.3389/fpsyg.2022.833395>.
22. de Oña J, de Oña R. Quality of service in public transport based on customer satisfaction surveys: a review and assessment of methodological approaches. *Transp Sci.* 2014;49(3):1–45. <https://doi.org/10.1287/trsc.2014.0544>.
23. Shek DTL. Is subjective outcome evaluation related to objective outcome evaluation? insights from a longitudinal study in Hong Kong. *J Pediatr Adolesc Gynecol.* 2014;27:S50–6. <https://doi.org/10.1016/j.jpag.2014.02.012>.
24. Chai W, Li X, Shek DTL. The effectiveness of a leadership subject using a hybrid teaching mode during the Pandemic: objective outcome and subjective outcome evaluation. *Int J Environ Res Public Health.* 2022;19(16):9809. <https://doi.org/10.3390/ijerph19169809>.
25. Spooren P, Mortelmans D, Denekens J. Student evaluation of teaching quality in higher education: development of an instrument based on 10 Likert-scales. *Assess Eval High Edu.* 2007;32(6):667–79. <https://doi.org/10.1080/02602930601117191>.
26. Shek DTL, Yang Z, Ma CMS, Chai CWY. Subjective outcome evaluation of service-learning by the service recipients: scale development, normative profiles and predictors. *Child Indic Res.* 2021;14:411–34. <https://doi.org/10.1007/s12187-020-09765-1>.
27. Yu L, Lin L, Shek TLD, Chai W. Students' perceived attributes and benefits of a leadership course: subjective outcome evaluation. *Res Soc Work Pract.* 2022;32(2):199–214. <https://doi.org/10.1177/10497315211042823>.
28. Zhu X, Shek DTL, Gong Z, Tang Y. Subjective outcome evaluation of a positive youth development program in mainland China during the COVID-19 pandemic. *Appl Res Qual Life.* 2025. <https://doi.org/10.1007/s11482-024-10416-8>.
29. Kirkpatrick DL. *Evaluating training program -The four levels.* San Francisco: Berrett-Koehler Publishers, Inc.; 1994.
30. Rönkkö M, Cho E. An updated guideline for assessing discriminant validity. *Organ Res Methods.* 2022;25(1):6–47. <https://doi.org/10.1177/1094428120968614>.
31. Lin L, Shek DTL, Li X. Who benefits and appreciates more? an evaluation of online service-learning projects in mainland China during the COVID-19 pandemic. *Appl Res Qual Life.* 2022. <https://doi.org/10.1007/s11482-022-10081-9>.
32. Mindrila D. Maximum likelihood (ML) and diagonally weighted least squares (DWLS) estimation procedures: a comparison of estimation bias with ordinal and multivariate non-normal data. *Int J Digit Soc.* 2010;1(1):60–6. <https://doi.org/10.20533/ijds.2040.2570.2010.0010>.
33. Li C-H. Confirmatory factor analysis with ordinal data: comparing robust maximum likelihood and diagonally weighted least squares. *Behav Res.* 2016;48:936–49. <https://doi.org/10.3758/s13428-015-0619-7>.
34. Bentler PM, Bonett DG. Significance tests and goodness of fit in the analysis of covariance structures. *Psychol Bull.* 1980;88(3):588–606. <https://doi.org/10.1037/0033-2909.88.3.588>.
35. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model.* 1999;6(1):1–55. <https://doi.org/10.1080/10705519909540118>.
36. Kline RB. *Principles and practice of structural equation modeling.* 3rd ed. New York: Guilford Press; 2011.
37. Browne MW, Cudeck R. Alternative ways of assessing model fit. *Social Methods Res.* 1992;21(2):230–58. <https://doi.org/10.1177/0049124192021002005>.
38. MacKenzie SB, Podsakoff PM, Podsakoff NP. Construct measurement and validation procedures in MIS and behavioral research: integrating new and existing techniques. *Mis Quart.* 2011;35(2):293–334. <https://doi.org/10.2307/23044045>.
39. Hu LT, Bentler PM. Evaluating model fit. In: Hoyle RH, editor. *Structural equation modeling: concepts, issues, and applications.* Thousand Oaks: Sage; 1995. pp. 76–99.
40. Raykov T. Behavioral scale reliability and measurement invariance evaluation using latent variable modeling. *Behav Ther.* 2004;35(2):299–331. [https://doi.org/10.1016/S0005-7894\(04\)80041-8](https://doi.org/10.1016/S0005-7894(04)80041-8).
41. Schmitt N. Uses and abuses of coefficient alpha. *Psychol Assess.* 1996;8(4):350–3. <https://doi.org/10.1037/1040-3590.8.4.350>.
42. Lin J-MS, Brimmer DJ, Maloney EM, Nyarko E, BeLue R, Reeves WC. Further validation of the Multidimensional Fatigue Inventory in a US adult population sample. *Popul Health Metr.* 2009;7(18):1–12. <https://doi.org/10.1186/1478-7954-7-18>.
43. Harding B, Tremblay C, Cousineau D. Standard errors: a review and evaluation of standard error estimators using Monte Carlo simulations. *Quant Method Psychol.* 2014;10(2):107–23. <https://doi.org/10.20982/tqmp.10.2.p107>.
44. Uttley J. Power analysis, sample size, and assessment of statistical assumptions -improving the evidential value of lighting research. *LEUKOS.* 2019;15(2–3):143–62. <https://doi.org/10.1080/15502724.2018.1533851>.
45. vanTassel-Baska J, Brown E. An analysis of stakeholder perceptions of gifted programs: a report card on gifted program performance. *Gifted Child Today.* 2022;45(3):160–75. <https://doi.org/10.1177/10762175221091859>.
46. Tommis S. Gifted education in the Hong Kong Special Administrative Region. *J Educ Gifted.* 2013;36(3):259–76. <https://doi.org/10.1177/0162353213492701>.
47. van der Meulen RT, van der Bruggen CO, Spilt JL, Verouden J, Berkhout M, Bögels SM. The pullout program day a week school for gifted children: effects on social-emotional and academic functioning. *Child Youth Care Forum.* 2013;43:287–314. <https://doi.org/10.1007/s10566-013-9239-5>.
48. Dimitriadis C. Provision for mathematically gifted children in primary schools: an investigation of four different methods of organizational provision. *Educ Rev.* 2012;64:241–60. <https://doi.org/10.1080/00131911.2011.598920>.
49. Chan DW. Characteristics and competencies of teachers of gifted learners: the Hong Kong teacher perspective. *Roeper Rev.* 2001;23(4):197–202. <https://doi.org/10.1080/02783190109554098>.
50. Chan LK, Chan DW, Sun X. University-based gifted programs for gifted and talented students in Hong Kong: practice and evaluation. *Gifted Educ Int.* 2020;36(2):90–107. <https://doi.org/10.1177/0261429420917858>.
51. Hall M, Jong MD, Steehouder M. Cultural differences and usability evaluation: individualistic and collectivistic participants compared. *Tech Communication.* 2004;51(4):489–503.
52. Noble H, Heale R. Triangulation in research, with examples. *Evid Based Nurs.* 2019;22(3):67–8. <https://doi.org/10.1136/ebnurs-2019-103145>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.