Int J Child Health Hum Dev 2016;9(2):173-183 ISSN: 1939-5965

Qualitative evaluation of a credit-bearing leadership subject in Hong Kong

Daniel TL Shek[1-5,*], PhD, FHKPS, BBS, SBS, JP, Hung-Kit Fok[1], PhD, Charles TL Leung[1], DSW Candidate, and Pecky PK Li[1], MSocSc

[1]Department of Applied Social Sciences, The Hong Kong Polytechnic University, Hong Kong, P.R. China
[2]Centre for Innovative Programmes for Adolescents and Families, The Hong Kong Polytechnic University, Hong Kong, P.R. China
[3]Kiang Wu Nursing College of Macau, Macau, P.R. China
[4]School of Social Development, East China Normal University, Shanghai, P.R. China
[5]University of Kentucky College of Medicine, Lexington, Kentucky, United States of America

Abstract

A subject entitled "Tomorrow's Leaders" aiming at promoting the holistic development of university students was offered to students at The Hong Kong Polytechnic University. Towards the end of the term, students were invited to give descriptors and metaphors about the subject. Based on the written narratives of 1,029 students, it was observed that most students used positive descriptors to describe the subject. Concerning the metaphors that could stand for the subject, most of the metaphors are positive in nature. Reliability analyses based on randomly selected coded responses showed that intra- and inter-rater reliabilities were high. In conjunction with other evaluation findings, the present study suggests that this subject was able to promote the holistic development in Chinese university students in Hong Kong.

Keywords: university students, Chinese adolescents; leadership; holistic development; qualitative evaluation; university students

Introduction

The quantitative-qualitative methodology debate appears to be everlasting in the field of evaluation (1-10). On the one hand, some researchers argued that quantitative and qualitative evaluation methods cannot be combined because they have different axiological and paradigmatic considerations (11, 12). According to Denzin and Lincoln (11), 'commensurability is an issue only when researchers want to "pick and choose" among the axioms of positivist and interpretivist models, because the axioms are contradictory and mutually exclusive' (p. 117). On the other hand, there are theorists arguing that both quantitative and qualitative methods are compatible for any evaluation research with reference to specific conditions of practice (13-16). Besides, the concept "methodological appropriateness" has been

* Correspondence: Professor Daniel TL Shek, PhD, FHKPS, BBS, SBS, JP, Associate Vice President (Undergraduate Programme) and Chair Professor of Applied Social Sciences, Department of Applied Social Sciences, The Hong Kong Polytechnic University, Room HJ407, Core H, Hung Hom, Hong Kong, P.R. China. E-mail: daniel.shek@polyu.edu.hk

coined (13, 14, 17, 18). According to Patton (17), methodological appropriateness was defined as "matching the evaluation design to the evaluation situation taking into account the priority questions and intended uses of primary intended users, the costs and benefits of alternative designs, the decisions that are to be made, the level of evidence necessary to support those decisions, ethical considerations, and utility" (p. 460). Patton (17) further proposed eight developments contributing to the feasibility of methodological appropriateness. For example, there is a growing consensus on interdisciplinary and multi-method approaches to evaluation.

With specific reference to the recent discussions of paradigms in the field evaluation, there are three main paradigms. They are post-positivism, constructivism and related perspectives, and pragmatism (17, 19-21). For example, Alkin (19) pointed out that the goal of post-positivistic research is trying to measure universal truth and its causal relationships among variables, although a perfect understanding of the truth cannot be achieved because of various observational errors. Constructionists, however, do not believe in one single truth; there are many truths that are relative to one's cultural and historical experiences. Relationships are also interdependent and thus no linear causal relationship can be found. Hence, observational biases need not be controlled but acknowledged. Moreover, the pursuit of local relevant findings is more appreciated than searching for universal truth. Pragmatist, like constructionists, also believe in no one single truth but one explanation of reality can be regarded "truer" than the other at a given timeframe; they also believe in causal relationships, like post-positivists, but in a more sophisticated manner (21, 22).

Besides the choice of paradigm in evaluation, it is also important to consider which method is suitable to be used under what conditions. In fact, a review of literature in educational evaluation has also shown that many approaches and models are available for various uses. First, House (23) proposed various evaluation models such as the utilitarian and intuitionist/pluralist perspectives. Second, Worthen and Sanders (24) classified six alternative approaches to educational evaluation. They are objectives-oriented, management-oriented, consumer-oriented, expertise-oriented, adversary-oriented, as well as naturalistic and participant-oriented approaches. Third, Scriven (25) suggested another six types of evaluation approaches, including strong decision support view, weak decision support view, relativistic view, rich description approach, social process school, and constructivist approach. Fourth, Stufflebeam, Madaus, and Kellaghan (26) analyzed twenty-two program evaluation approaches to identify which approach is the most worthy one to be applied under certain conditions. Fifth, Cousins and Ryan (27) suggested four ongoing issues for educational evaluation, including evidence-based policy and programming, performance measurement, auditing and monitoring, learning and discovery-oriented evaluation, and values-oriented evaluation.

In the context of higher education, many approaches have similarly been used to assess the effectiveness of courses and programs. Among the various approaches that have been adopted, it is common to conduct mandatory course evaluation (i.e., subjective outcome evaluation) at the end of the semester (28). As a means of obtaining information about the strengths and weaknesses of a course and to make instructional and administrative decision (29), course evaluation usually takes the form of a series of ubiquitous Likert-scales (28, 30, 31). Although course evaluation is an efficient way to gauge the views of the students, it might not be able to reflect the complexity of the subjective learning experiences (32). Starr-Glass (28) regarded course evaluation based on the questionnaire approach "fail to capture the quintessential dimensions of learning engagement" (p. 198). Parlett and Hamilton (33) commented many formal evaluations present "an emaciated and artificial picture of real education life" (p. 16). In order to address these criticisms, it is argued that qualitative evaluation methods could give some additional information on the merits and effectiveness of a course.

There are many qualitative evaluation approaches and methods (13, 14, 34, 35), such as focus groups, interviews, and written reflections. In this study, qualitative data in the form of written descriptors and metaphors were collected to understand the perceived benefits of a subject. The concept of metaphor has been emphasized in the field of educational evaluation (14, 36-39). Metaphors are used in everyday communication and they can be regarded as good

representations of realities (40). Metaphor has been conceived as a fundamental vehicle that people use to understand, express, construct, and organize their world (39-42). It provides insights for new perspectives and innovation (39), helps people to explain unknown experiences by known one (42), serves as a convenient way to communicate a complex idea (43), describes important features of a complex array variables in a simple form (44), transmits a complete story using simple images (45), and offers dynamic and dramatic views beyond the surface of things and reveals their deeper significance (46).

In term of conceptualization of metaphor as an analogy, metaphor is defined as an analogy that imaginatively identifies one object with another (47) and as an analog device that serves as a means for framing and defining experiences (48). Yob (49) further explained that metaphor is employed when one wants to explore and understand something esoteric, abstract, novel, or highly speculative.

In the domain of education research, Martinez, Sauleda and Huber (50) argued that teaching and learning could be addressed from three major perspectives. The first one is the behaviorist/ empiricist perspective. This perspective regards the learner as passive and knowledge is developed by forming associations and by subdivision of learning tasks into small and logically sequenced components. Under this perspective, teachers may be treated as transmitters of information whereas learners as passive recipients in metaphors. The second one is the cognitive/constructivist perspective. This perspective regards development of knowledge as actively constructed by the learner though transforming old schema into new one. Under this perspective, teachers may be treated as facilitators of information whereas learners as active agents of learning process in metaphors. The third one is the situational/socio-historical perspective. Under this perspective, learning is situated in the social context in which it is constructed. Knowledge is a by-product of the activity, context, and culture in which it is used. Classifying metaphors of teaching and learning into three major dimensions provides a basic conceptual framework to understand the educational beliefs of the teachers and students.

Concerning the meaning of metaphor in the field of education evaluation, Madaus and Kellaghan (51) argued that metaphors "influence the way we understand and talk about education; they create mind-sets and influence behavior towards school and teachers; they also influence the kinds of questions we ask about educational programs." Results of related research indicate that metaphor is able to illustrate diversified and authentic understanding of the respondents (28, 32, 52, 53). Alvesson (54) further suggested that a sophisticated use of metaphor can "facilitates offering various comprehensive images of research, thus reducing the risk of latching on to a one-sided and favourite conception." However, Alvesson (54) also stated that "we can think in contradiction to our favored metaphors and not just deviate from the thinking they encourage in those exceptional cases where they clearly don't work." This reminder of using metaphor is also shared by other scholars (55, 56).

Nevertheless, the use of metaphor in education contexts does contain drawbacks (57). First, metaphors have a constrained conceptual framework because the assumptions and predispositions reflected by the metaphors can only relate to the phenomena the author intends to cover (56). Second, metaphors may create distortions because metaphors create "a way of seeing" and "ways of not seeing" at the same time (55). Despite these limitations, research shows that there are several merits of using metaphors in education contexts (57). First, metaphors provide coherent and internal consistent information that give insight into implicit ideas. Second, metaphors are evocative and stimulating for people to tease out connections that might not be made use of by direct questions (57). Third, metaphors synthesize a large amount of knowledge about teachers, learners, and teaching methods (58). Fourth, metaphors serve as tools for teachers to gain distances from their own practices and to act as external observers reviewing their teaching. Fifth, metaphors make implicit knowledge explicit through reflection on and representation of the concepts. Sixth, metaphors give a language that can connect theories and practices. Seventh, metaphors empower teachers to examine their own assumptions and to explore hidden intellectual avenues (59). Finally, metaphors reflect

upon alternative teaching practices and theoretical frameworks (57).

Given the fact that metaphors and related concepts like descriptors have many strengths, metaphors play a central role in conceptualizing and reflecting upon the nature of teaching and learning (57) and they have been used as tools to evaluate education programs. For instance, to evaluate a distance-learning course in business and economics at the undergraduate level, Starr-Glass (28) asked students to provide semantic indicators and metaphors and found that the course was promoting independent thought, consideration, and reflection. In addition, with the use of semantic indicators and metaphors, Starr-Glass (28) identified "a great of ostensibly significant and authentic emotional engagement with the course" (p. 205). Thomas and Beauchamp (32) used metaphors in semi-structured interviews to explore the learning experiences of newly graduated teachers in term of their professional identity. Newly graduated teachers were interviewed immediately after graduation of their teacher education programs and after their first year of teaching. The participants were asked to create their own metaphors on teaching experiences and teaching self. The metaphors were coded using N-vivo. The findings indicated that new teachers struggled to develop their professional identity during their first year, and that this development process was gradual, complex, and often problematic. Research also examined how metaphor revealed changes of pre-service elementary teachers' belief about learning and teaching. The factors contributed to the changes of pre-service elementary teachers' belief were also examined (57).

Finally, Levin and Wagner (52) also used metaphors to explore eighth grade students' view on writing in science classrooms. They divided 97 eighth graders into intervention classes and comparative classes. In the intervention classes, students were assigned to complete informal reflective writing tasks whereas the students in comparative classes were not required to write anything. At the end of the teaching units, all students were required to submit two final reflective writings. Levin and Wagner (52) analyzed the students' writing and found that students in the intervention group expressed emotional dimension less than students in the comparative group. Students in the intervention group expressed the social dimension more than students in the comparative group. Levin and Wagner [52] further analyzed that content of the reflective writing. The results showed that writing-to-learn task, feedback and reflective writing greatly influenced students' view on writing.

Despite research utilizing metaphors and related descriptors are available in education evaluation research, studies focusing on learning experiences from students' perspective at university education setting was limited. A computer search of PsycINFO in March 2014 using the search terms of "metaphor," "university education," and "evaluation" showed that there were 52 citations. A similar survey of Social Sciences Citation Index using the same search terms only located 3 citations. If adding "descriptor" as a search term, no citation can be found. These figures imply that there are only few studies using metaphors and descriptors to understand the subjective learning experiences from students' perspectives, especially in the Chinese context (60). To fill this research gap, the present study intended to explore the views of university students on a subject on leadership and intrapersonal development using metaphors and descriptors.

To promote holistic development of university students and address community concern about young people (61), a general university requirements subject titled "Tomorrow's Leaders" was offered at The Hong Kong Polytechnic University to cultivate various intra and inter-personal competences of students. Through lectures, experiential learning, group presentation and individual assignments, students were nurtured on their self-understanding, personality, emotional competence, cognitive competence, resilience, spirituality, social competence, moral competence, positive identity, interpersonal communication, conflict resolution, team-building, and relationship-building.

In the piloting phase, research indicated the effectiveness of the course by several evaluation strategies. These included: a) pre-test/post-test questionnaire using the Chinese Positive Youth Development Scale to examine changes in intra- and interpersonal qualities in students (62), b) subjective outcome evaluation at the end of each lecture, c) post-course subjective outcome evaluation conducted at the end of the whole course (63), d) process evaluation via systematic observation (64), e) focus

group interview for students' view on their learning experiences (65), f) descriptors and metaphor about the students' feelings and perception of the course in a reflection sheet. Based on the data collected in the formal implementation phase, the objective of this study was to evaluate the subject based on the descriptors and metaphors written by the students.

Methods

Sixteen classes of students took this course, with a total enrollment of 1,029 students (57 in Class A, 52 in Class B, 58 in Class C, 70 in Class D, 59 in Class E, 55 in Class F, 83 in Class G, 66 in Class H, 84 in Class I, 73 in Class J, 91 in Class K, 72 in Class L, 56 in Class M, 58 in Class N, 52 in Class O and 43 in Class P). At the end of the course, the personal reflection forms were given to the students in which they were invited to fill in two parts. In the first part, students were invited to use three words or phrases to describe their feelings, perceptions, and experiences about taking this subject (i.e., descriptors of this course). In the second part, students were invited to think about a metaphor which could stand for the course and gave some explanations about the meaning of the metaphor. The metaphor may be an object, an event, or a state (e.g., a wonderful journey, a seed and a book).

Data analyses

A total of 633 personal reflection forms were collected from the students. Their responses in the reflection form were entered in the computer for analyses. Four researchers were involved in the data analyses process. Two research assistants were responsible for the "descriptor" part and the other two were responsible for the "metaphor" part. A meaningful unit was used as the basic unit of analysis. For instance, the statement that a course was "wonderful and relaxing" was broken down into two meaningful units, namely, "wonderful" and "relaxing." On the other hand, descriptions with the same meaning (e.g., "relaxed" and "relaxing") were grouped into the same attribute category. Regarding the positivity of the codes, four possible codes were involved, including positive code, negative code, neutral code and undecided code. A total of 1,904 descriptors and 613 metaphors were coded and analyzed.

According to Miles and Huberman (66), check-coding facilitated definitional clarity and it also served as a good reliability check. Check-coding becomes sharper when two researchers code the same data set and discuss their initial difficulties. To enhance the reliability of the coding on the positivity nature of the raw codes, both intra- and inter-rater reliability were conducted. Regarding the first part (i.e., descriptors), 40 randomly selected descriptors were re-coded among the two researchers to check for the intra-rater reliability. Regarding the inter-rater reliability, 40 randomly selected descriptors were also re-coded among the other two researchers who did not involve in the initial coding process.

Regarding the second part (i.e., metaphors), 40 randomly selected metaphors were re-coded among the two researchers to check for the intra-rater reliability. With a view to check for the inter-rater reliability, 40 randomly selected descriptors were also recoded among the two researchers who were not involved in the initial coding process. The raw data files and steps involved in the development of coding system were properly documented and systematically organized.

Results

Qualitative findings on two areas are presented as follows: (a) descriptors that were used by the participants to describe the program; and (b) metaphors (objects, occasions, feelings) that were used by the participants to portray the program. The descriptors that the participants used to describe the program are presented in Table 1. There were 1,881 raw descriptors in total, which could be further classified into different classifications. Among all of the descriptors, 1,745 (92.8%) of them were positive responses, 67 (3.6%) of them were negative responses (see Table 2), while 46 (2.5%) of them were classified as neutral responses and 23 (1.2%) of them were undecided. It was found that participants used "boring," "demanding," "long lesson," "useless," "abstract" to describe most of the negative responses,

while the remaining responses were classified as other negative responses. Eighty raw descriptors were randomly selected for the calculation of the reliability tests. The intra-rater reliability was 100% whilst the inter-rater reliability was 95%.

Table 1. Categorization of descriptors to describe the course

Descriptors	Nature of the response				
	Positive	Neutral	Negative	Undecided	Total
Comprehensive	44				44
Creative	13				13
Enjoyable	63				63
Funny	70				70
Good	51				51
Happy	44				44
Inspiring	204				204
Interactive	58				58
Interesting	237				237
Knowledgeable	9				9
Leadership	7				7
Meaningful	142				142
Positive	19				19
Practical	4				4
Reflective	92				92
Relaxing	140				140
Self-understanding	15				15
Touching	7				7
Unique	4				4
Useful	286				286
Warm	14				14
Well organized	4				4
Co-operation	23				23
Self-Improvement	17				17
Teaching Staff attributes	20				20
Other positive descriptors (e.g., I'm loving it, time slides when you enjoy yourself)	158				158
Memorable		19			19
Other neutral descriptors (e.g., games, new experience to me)		27			27
Abstract			3		3
Boring			17		17
Demanding			10		10
Long lesson			6		6
Useless			5		5
Other negative descriptors (e.g. confused, tired)			26		26
Academic				14	14
Creative				1	1
Unexpected				8	8
Total count, n	1745	46	67	23	1881
Total count, %	92.77	2.45	3.56	1.22	100

Table 2. Descriptors coded as negative responses

Negative descriptors	n
Boring	17
Demanding	10
Long lesson	6
Useless	5
Abstract	3
Other negative descriptors	26
Total	67

Table 3. Categorization of the metaphors to describe the program

Metaphor	Nature of metaphor					Number of codes derived from the metaphor and its nature				
	Positive	Neutral	Negative	Undecided	Total	Positive	Neutral	Negative	Undecided	Total
Air and earth	5				5	6				6
Apple	6				6	11				11
Beacon	7	2			9	12				12
Book	18				18	21	1	2		24
Box	1			1	2	1			1	2
Buffet	9			1	10	11	1	1		13
Chinese medicine	4				4	6		2		8
Compass	30				30	39				39
Cooperation	1			1	2	1			1	2
Dictionary & Wikipedia	4	1			5	4	1			5
Garden	5				5	5			1	6
Journey	30	1	1		32	45	3	2		50
Lego	4				4	4				4
Library	4				4	7	1			8
Light	29				29	46				46
Map	7				7	11				11
Microscope	7				7	9				9
Mirror	36				36	50				50
Other metaphors	277	20	9		306	437	28	31	2	498
Pizza	4				4	7				7
Planting	6				6	5				5
Rainbow	9				9	13				13
Sea and sky	5				5	5				5
Seed	7	1			8	8	1			9
Snow ball	4				4	5	1			6
Star	4			1	5	4			1	5
Sweets	15	1		1	17	24		3	1	28
Tree	6				6	13				13
Water and sun	25	1			26	45	2			47
Annoy orange				1	1				1	1
Rain in spring and snow in winter				1	1				1	1
Total count, n	569	27	10	7	613	855	39	41	9	944
Total count, %	92.82	4.4	1.63	1.14	100	90.57	4.13	4.34	0.95	100

Table 4. Illustrations of metaphors coded as negative

1. But some are boring
2. Boring
3. But could not go really deep
4. Bitter
5. Process is hard
6. Theories aren't attractive
7. Boring and dull
8. Bitter at first
9. Boring at first
10. Demanding workload
11. Don't find my way in the first lecture.
12. Intangible
13. Annoying
14. Appearance is bad
15. Bored
16. Boring lectures
17. But it is not yummy
18. Difficult in project part
19. Difficult to use
20. Do not want to do it at the beginning
21. I did not pick up anything
22. I feel tired, just like what coffee does on people
23. It is not an interesting subject
24. Long lecture hour
25. Many elements but they are difficult to learn/ understand
26. Mysterious
27. Not attractive
28. Not easy to learn those concepts, just like it's not easy to search for the eggs.
29. Not really helpful to me
30. Not useful to me
31. Quite confusing
32. Running time is long
33. She is which I wish to understand, but I can't understand
34. Short time to master them
35. Students did have chances to do reflection in class but they simply didn't have the initiative to do so
36. The subject on a whole is quite time-consuming with the assignment and projects
37. Topics are not connected
38. Useless

The categorization of metaphors written by the participants is shown in Table 3. There were 613 raw metaphors which could be further categorized into different categories. Among the responses, 569 (92.8%) of them were positive responses while 10 (1.6%) of them were negative responses. The description for the negative responses can be seen in Table 4. Results revealed that participants used different expressions to describe the program, with most using "mirror" as the metaphor. Some participants used "journey," "compass," "light" and "water and sun" to describe the program. Eighty raw metaphors were randomly selected for the calculation of intra and inter-rater reliability of the coding. The related values were 95% and 92.5%, respectively.

Discussion

Based on the qualitative data on the descriptors and metaphors collected from students, the present study investigated perceptions of students of "Tomorrow's

Leaders," a credit-bearing course enhancing holistic development at The Hong Kong Polytechnic University, in the full implementation stage. There were several distinctive features of the present study. First, the present study provided a qualitative evaluation on the subjective learning experiences of the students based on metaphors and descriptors. As there are few related studies in this area, this study fills the research gap in this area. Second, this study demonstrated how qualitative data of the metaphors and descriptors from students supplemented quantitative course evaluation using ubiquitous Likert-scales. Unlike numbers and statistics, the metaphors and descriptors showed complexity of the subjective learning experiences of the students (32). Third, a large sample was used in this study. This is unique because the sample size in qualitative studies is usually small.

This study provided evidence supporting the positive impacts and effectiveness of "Tomorrow's Leaders" based on qualitative data on the subjective experiences of the students. Two salient conclusions can be drawn from the present study. First, students perceived the course positively. Roughly nine-tenth of the responses in the descriptors and metaphors were positive (see Tables 1 and 3). For example, the students described the course as "interesting," inspiring," and "useful" according to the descriptors collected (see Table 3). Most of the negative responses were related to the views that the course was boring and academically demanding. Although the negative responses accounted only 10% of the responses, this information was useful for further improvements of the course. Another observation was some negative responses were contradictory in nature. For instance, some students reported that the lecture time was too long whereas some regarded that the lecture was too short for them to master the knowledge.

The positive feedbacks obtained in this study are consistent with other evaluation studies based on different methods (60, 62-65, 67-73). For instance, objective outcome evaluation findings showed that students had positive changes in intra- and inter-personal competences after taking the course (62). In addition, both subjective outcome evaluation at the end of each lecture and post-course subjective outcome evaluation showed that students had positive perception of the program, implementers, and effectiveness (63, 71). Process evaluation via systematic observation also showed that the program had high program adherence (68). In sum, research findings in different studies suggested that the course brought positive impacts on students for their holistic development and students had positive views on the course.

Nevertheless, although positive feedbacks were documented among focus groups, alternative explanations should be acknowledged (74). First, positive responses from students might be due to demand characteristics. However, this explanation was not plausible because students freely gave their responses and negative comments were noted. In addition, the identity of the students was not revealed in their submission. Second, favorable results might be explained by researchers' preoccupations about positive evaluation outcomes. However, intra-and inter-rater reliability as well as disciplined data analyses were performed to reduce possible biases in data collection and analyses.

While the findings are positive, it should be noted that the present study faced several limitations. First, metaphors and descriptors from students were collected towards the end of the course. It would be illuminating if data could be collected at different time points (e.g., mid-term evaluation). Second, in addition to metaphors and descriptors, in-depth individual interviews of both students and teachers could be attempted. This can allow researchers to understand the subjective experiences in different stakeholders. Finally, due to practical and time constraint, peer-checking and member checking were not carried out. It would be desirable to add such feedback mechanisms in the future. Despite these limitations, the present study provides qualitative findings to support the positive features and effectiveness of the course "tomorrow's leaders" and its effectiveness from the perspective of the students.

References

[1] Bredo E. Comments on Howe: getting over the methodology wars. Educ Res 2009;38:441-8.

[2] Howe KR. Two dogmas of educational research. Educ Res 1985;14:10-8.

[3] Howe KR. Against the quantitative-qualitative incompatibility thesis or dogmas die hard. Educ Res 1988;17:10-6.
[4] Howe KR. Closing methodological divides: toward democratic educational research. Dordrecht, Netherlands: Kluwer, 2003.
[5] Howe KR. Isolating science from the humanities: the third dogma of educational research. Qual Inq 2009;15:766-84.
[6] Johnson RB. Comments on Howe: toward a more inclusive "scientific research in education". Educ Res 2009;38:449-57.
[7] Salomon G. Transcending the qualitative-quantitative debate: the analytic and systemic approaches to educational research. Educ Res 1991;20:10-8.
[8] Smith JK. Quantitative versus qualitative research: an attempt to clarify the issue. Educ Res 1983;12:6-13.
[9] Smith JK, Heshusius L. Closing down the conversation: the end of the quantitative-qualitative debate among educational inquirers. Educ Res 1986;15:4-12.
[10] Tillman LC. Comments on Howe: the never-ending education science debate: I'm ready to move on. Educ Res 2009;38:458-62.
[11] Denzin NK, Lincoln YS. The Sage handbook of qualitative research, 4th ed. Thousand Oaks, CA: Sage, 2011.
[12] Guba EG, Lincoln YS. Competing paradigms in qualitative research. In: Denzin NK, Lincoln YS, editors. The handbook of qualitative research. Thousand Oaks, CA: Sage, 1994:105-17.
[13] Patton MQ. Qualitative evaluation and research methods, 2nd ed. Newbury Park, CA: Sage, 1990.
[14] Patton MQ. Teaching and training with metaphors. Am J Eval 2002;23:93-8.
[15] Reichardt CS, Cook TD. Beyond qualitative versus quantitative methods. In: Cook TD, Reichardt CS, editors. Qualitative and quantitative methods. Beverly Hills, CA: Sage, 1979:7-32.
[16] Reichardt CS, Rallis SF. The qualitative-quantitative debate: new perspectives. New Dir Program Eval 1994;61:1-98.
[17] Patton MQ. Utilization-focused evaluation, 4th ed. Thousand Oaks, CA: Sage, 2008
[18] Donaldson SI, Patton MQ, Fetterman DM, Scriven M. The 2009 claremont debates: the promise and pitfalls of utilization-focused and empowerment evaluation. J Multidisc Eval 2010;6:15-57.
[19] Alkin MC. Evaluation roots: a wider perspective of theorists' view and influences. Thousand Oaks, CA: Sage, 2013.
[20] Mertens DM. Transformative mixed methods research. Qual Inq 2010;16:469-74.
[21] Patton MQ. Developmental evaluation: applying complexity concepts to enhance innovation and use. New York: Guilford, 2011.
[22] Funnell SF, Rogers PJ. Purposeful program theory: effective use of theories of change and logic models. San Francisco, CA: Jossey-Bass, 2011.
[23] House ER. Assumptions underlying evaluation models. Educ Res 1978;7:4-12.
[24] Worthen BR, Sanders JR. Educational evaluation: alternative approaches and practical guidelines. New York: Longman, 1987.
[25] Scriven M. Evaluation as a discipline. Stud Educ Eval 1994;20:147-66.
[26] Stufflebeam DL, Madaus GF, Kellaghan T. Evaluation models: viewpoints on educational and human services evaluation. Boston, MA: Kluwer, 2000.
[27] Cousins JB, Ryan KE. Dilemmas for educational evaluation in a globalized society. In: Ryan KE, Cousins JB, editors. The Sage international handbook of educational evaluation. Thousand Oaks, CA: Sage, 2009:539-54.
[28] Starr-Glass D. Metaphor and maps in evaluation. Assess Eval Higher Educ 2005;30:195-207.
[29] Worthen BR, Sanders JR. Educational evaluation: alternative approaches and practical guidelines. New York: Longman, 1987.
[30] Marsh HW, Ginns P, Morin AJS, Nagengast B, Martin AJ. Use of student ratings to benchmark universities: multilevel modeling of responses to the Australian Course Experience Questionnaire (CEQ). J Educ Psychol 2011:103:733-48.
[31] Marsh HW, Lüdtke O, Nagengast B, Trautwein U, Morin AJS, Abduljabbar AS, et al. Classroom climate and contextual effects: conceptual and methodological issues in the evaluation of group-level effects. Educ Psychol 2012;47:106-24.
[32] Thomas L, Beauchamp C. Understanding new teachers' professional identities through metaphor. Teach Teach Educ 2011;27:762-69.
[33] Parlett M, Hamilton D. Illumitative evaluation: a new approach to the study of innovatory programs. University Centre for Research in Educational Science, UK: University of Edinburgh, 1972.
[34] Fetterman DM. Qualitative approaches to evaluation in education: the silent scientific revolution. NY: Praeger, 1988.
[35] Patton MQ. Qualitative evaluation methods. Beverly Hills, CA: Sage, 1980.
[36] House ER. New directions in educational evaluation. London, Philadelphia, PA: Falmer, 1986.
[37] Kaminsky A. Beyond the literal: metaphors and why they matter. New Dir Eval, 2000;86:69-80.
[38] Smith NL. Metaphors for evaluation: source of new methods. Beverly Hills, CA: Sage, 1981.
[39] Smith NL. Mining metaphors for methods of practice. In: Fetterman DM, editor. Qualitative approaches to evaluation in education: the silent scientific revolution. New York: Praeger, 1988.

[40] Lakoff G, Johnson M. Metaphors we live by: with a new afterword. Chicago, IL: University of Chicago, 2003.
[41] Kliebard HM. Curriculum theory as metaphor. Theory Pract 1982;21:11-17.
[42] Munby H. Metaphor and teacher's knowledge. Res Teach Engl 1987;21:377-97.
[43] Crider C, Cirillo L. Systems of interpretation and the function of metaphor. J Theory Soc Behav 1991;21:171-95.
[44] Morgan G. Paradigms, metaphors, and puzzle solving in organization theory. Adm Sci Q 1980;25:605-22.
[45] Dickmeyer N. Metaphor, model, and theory in education research. Teach Coll Rec 1989;91:151-60.
[46] Marchant GJ. A teacher is like a…: using simile lists to explore personal metaphors. Lang Educ 1992;6:33-45.
[47] Holman CH. A handbook to literature. Indianapolis, IN: Bobbs-Merrill, 1980.
[48] Shaw DM, Mahlios M. Pre-service teachers' metaphors of teaching and literacy. Read Psychol 2008;29:31-60.
[49] Yob IM. Thinking constructively with metaphors. Stud Philos Educ 2003;22:127-38.
[50] Martinez MA, Sauleda N, Huber GL. Metaphors as blueprints for thinking about teaching and learning. Teach Teach Educ 2001;17:965-77.
[51] Madaus GF, Kellaghan T. Models, metaphors, and definitions of evaluation. In: Stufflebean DL, Madaus GF, Kellaghan T, eds. Evaluation models: viewpoints on educational and human services evaluation. Boston, MA: Kluwer, 2000:19-32.
[52] Levin T, Wagner T. In their own words: understanding student conceptions of writing through their spontaneous metaphors in the science classrooms. Instr Sci 2006;34:227-78.
[53] Saban A. Prospective teachers' metaphorical conceptualizations of learner. Teach Teach Educ 2010; 26:290-305.
[54] Alvesson M. Interpreting interviews. London: Sage, 2011.
[55] Morgan G. Images of organization. Beverley Hills, CA: Sage, 1986.
[56] Phillips DC. Philosophical perspectives. In: Berliner DC, Calfee RC, editors. Handbook of educational psychology. New York: Simon Schuster Macmillan, 1996:1005-19.
[57] Leavy AM, McSorley FA, Boté LA. An examination of what metaphor construction reveals about the evolution of preservice teachers' beliefs about teaching and learning. Teach Teach Educ 2007;23:1217-33.
[58] Calderhead J, Robson M. Images of teaching: Student teachers' early conceptions of classroom practice. Teach Teach Educ 1991;7:1-8.
[59] Connelly FM, Clandinin JD, He MF. Teachers' personal practical knowledge on the professional knowledge landscape. Teach Teach Educ 1997;13:665-74.
[60] Shek DTL, Sun RCF. Qualitative evaluation of a positive youth development course in a university setting in Hong Kong. Int J Disabil Hum Dev 2012;11:243-8.
[61] Shek DTL. Nurturing holistic development of university students in Hong Kong: where are we and where should we go? ScientificWorldJournal 2010;10:563-75.
[62] Shek DTL, Sun RCF. Promoting psychosocial competencies in university students: evaluation based on a one group pretest-posttest design. Int J Disabil Hum Dev 2012;11:229-34.
[63] Shek DTL, Sun RCF. Post-lecture evaluation of a university course on leadership and intrapersonal development. Int J Disabil Hum Dev 2013;12:185-91.
[64] Shek DTL, Sun RCF. Process evaluation of a leadership and intrapersonal development subject for university students. Int J Disabil Hum Dev 2013;12: 203-11.
[65] Shek DTL, Sun RCF. Focus group evaluation of a positive youth development course in a university in Hong Kong. Int J Disabil Hum Dev 2012;11:249-54.
[66] Miles MB, Huberman AM. Qualitative data analysis: an expanded sourcebook. Thousand Oaks, CA: Sage, 1994.
[67] Shek DTL, Sun RCF. Promoting leadership and intrapersonal competence in university students: what can we learn from Hong Kong? Int J Disabil Hum Dev 2012;11:221-28.
[68] Shek DTL, Sun RCF. Process evaluation of a positive youth development course in a university setting in Hong Kong. Int J Disabil Hum Dev 2012;11:235-41.
[69] Shek DTL, Sun RCF, Chui YH, Lit SW, Yuen WW, Chung YYH, Ngai SW. Development and evaluation of a positive youth development course for university students in Hong Kong. ScientificWorldJournal 2012:http://dx.doi.org/10.1100/2012/263731.
[70] Shek DTL. Post-lecture evaluation of a positive youth development subject for university students in Hong Kong. ScientificWorldJournal 2012:http://dx.doi.org/10.1100/2012/934679.
[71] Shek DTL, Sun RCF. Post-course subjective outcome evaluation of a course promoting leadership and intrapersonal development in university students in Hong Kong. Int J Disabil Hum Dev 2013;12:193-201.
[72] Shek DTL. Reflections of Chinese students on a university subject on leadership and intrapersonal development. Int J Disabil Hum Dev 2013;12:213-9.
[73] Shek DTL, Sun RCF, Tsien-Wong TB, Cheng CT, Yan HR. Objective outcome evaluation of a leadership and intrapersonal development subject for university students. Int J Disabil Hum Dev 2013;12:221-7.
[74] Shek DTL, Tang VM, Han XY. Quality of qualitative evaluation studies in the social work literature: evidence that constitutes a wakeup call. Res Soc Work Pract 2005;15:180-94.

Submitted: September 18, 2014. *Revised:* September 30, 2014. *Accepted:* October 13, 2014.