Issues in Measuring Generic Skills Using Self-Administered Questionnaire in a Community College in Hong Kong

Wai Tung Ho, Joseph Chi Ho So

The Hong Kong Polytechnic University, Hong Kong Community College, Hong Kong
Email: ccwtho@hkcc-polyu.edu.hk, ccso@hkcc-polyu.edu.hk

Abstract

In this study, we evaluate the reliability and validity of Self-Assessment of All-Round Development (SAARD) Questionnaire developed by The Hong Kong Polytechnic University when applied to two cohorts of students in Hong Kong Community College. Despite of the drastic changes in the educational system in Hong Kong, the SAARD questionnaire can still maintain a similar level of reliability as compared to the pilot study carried out a decade ago. However, this study reveals that there are too many items in the questionnaire and fatigue effect is affecting the reliability of items near the end of the survey. In addition, three of the items in the questionnaire have extremely large variations, implying potential problems in the wording of questions corresponding to these items. Remedies are also proposed in this study to alleviate the problems identified in the SAARD questionnaire.

1. Introduction

With the explosion of association degree programs in Hong Kong in the last decade, educators and administrators of the sector are eager to assess the effectiveness of associate degree education in Hong Kong. The success of these programs is not only being evaluated based on the academic achievements of the students but also the generic skills they developed in their college years. These generic skills are often essential to a student's future development beyond the classroom. Many researches were devoted to the understanding of generic skills and factors in driving the development of generic skills [1]-[7].

In this article, we will look at the generic skill assessment approaches employed by Hong Kong Community College (HKCC), a subsidiary of The Hong Kong Polytechnic University (PolyU). The college is one of the largest associate degree institute in Hong Kong. In HKCC, generic skills are integral part of learning objectives of each program. Direct and indirect measures are employed to assess graduates' academic competency as well as their generic skills. Direct measures used include measures embedded in assignments, tests, projects and presentations. While indirect measures are usually used to collect assessments from the perspective of the college's articulation partners, such as universities and employers.

One direct measure used in assessing students' generic skills is the Self-Assessment of All-Round Development (SAARD) questionnaire developed by PolyU in 2006 (Fung et al. [8] [9]). In HKCC, samples of freshmen and graduates are asked to complete the SAARD questionnaire. The results enable the college to evaluate the effectiveness of their programs and student development activities in promoting generic skills. Due to its critical role, it is important to ensure the reliability of SAARD questionnaire. The reliability and validity of the SAARD questionnaire were evaluated in depth by Fung et al. [8] [9] in their pilot study conducted in 2006. The questionnaire was found to be effective in measuring various generic skills of university students. However, the education sector in Hong Kong has gone through a number of significant changes in the last decade. That includes the rapid expansion of the associate degree sector, adoption of 4-year curriculum in the university and the transition to Hong Kong Diploma of Education (DSE) in secondary schools. Since all these changes lead to a change in the student population as well as their skill sets when entering college. It becomes an urgent matter to assess the reliability of SAARD which is developed almost a decade ago and its applicability in assessing generic skills of today's students.

2. About the SAARD Questionnaire

The SAARD questionnaire has 56 questions covering 14 generic skills: communications; creative thinking; critical thinking; cultural appreciation; entrepreneurship; EQ & psychological wellness; global outlook; healthy lifestyle; interpersonal effectiveness; leadership; lifelong learning; problem solving; social and national responsibility; and teamwork. Each generic skill is covered by 4 questions using a 7-point Likert scale (1 point = not well, 7 point = very well).

In HKCC, the survey is carried out on each cohort of students in two phases. In the first phase, a sample of students will be selected to complete the questionnaire at "entry", that is their first semester of study in the college. In the second phase, another sample of students will be selected at "exit", that is their last semester of study in the college. The change in generic skills over the two-year period of their studies in the college is of interest to various parties in the college. To improve the response rate, the survey is typically carried out during class hours. The collected data will be merged with administrative data for further statistical analysis. Analyses of the data collected can be found in (So et al. [10]).

3. Potential Issues of SAARD

As described above, SAARD was originally developed in 2006 for use in PolyU covering students in higher diploma to postgraduate programs. The students in HKCC are similar in level to the higher diploma students in PolyU. Despite of the similarity, it is reasonable to question if the SAARD is still a good instrument for measuring generic skills consider the drastic changes in the education system in secondary and tertiary levels. In particular, we will look at the student cohorts admitted to HKCC in 2011/12 academic year and 2013/14 academic year. These two cohorts represent the student populations who are under A-Level curriculum and DSE curriculum respectively. Using the data collected from the two cohorts mentioned above, we will study the reliability and validity of the SAARD survey by comparing to the PolyU pilot study in 2006.

The second issue of SAARD is the length of the questionnaire. In order to cover 14 different generic skills, 4 questions are asked for each of the generic skill, giving us a total of 56 questions. It is a concern that the quality of data collected near the end of the questionnaire is affected by fatigue effect. From the administration point of view, a long questionnaire is also undesirable because SAARD surveys are normally done during class hours. A lengthy questionnaire will undoubtedly cause unnecessary interruption to teaching.

4. About the Samples and the Cohorts

The data being used in this study were collected from two cohorts of students entering the college in 2010 and

2011 academic years. The choice of these cohorts instead of newer cohorts enables us to validate published results published by PolyU in 2006. Due to the secondary school reforms, later student cohorts might behave differently when completing the SAARD survey. The analysis of generic skills of later student cohorts will be done in a separate study (**Table 1**).

For each of the two cohorts in this study, random samples of students were asked to complete the SAARD questionnaire at the beginning of their first semester in the college (*i.e.* entry survey). At the end of the last semester of their studies, another random sample of students was asked to complete the SAARD questionnaire.

5. Reliability Analysis of SAARD Surveys

To assess the quality of the measurements obtained in our survey, reliability statistics of our studies are compared to the results of the PolyU pilot conducted in 2006. **Table 2** and **Table 3** show the Cronbach alphas and mean inter-item correlations for the entire questionnaire and the 14 scales being measured.

The reliability statistics of the two HKCC cohorts are mostly in line with the PolyU pilot. The smaller sample sizes at "Exit" have only caused a mild degradation in the reliability statistics. Among all scales, Social and National Responsibility has the lowest Cronbach alpha and mean inter-item correlations across the four samples. This indicates potential issues of the four items used in this scale. The exact wordings of the four items under Social and National Responsibility are:

Table 1. Sample sizes of 2010 and 2011 cohorts.

Sample Size	2010/11 Cohort	2011/12 Cohort
Entry (beginning of the first semester)	1377	933
Exit (end of the last semester)	564	347

Table 2. Reliability of SAARD questionnaire 2010/11 cohort.

Scale	No. of Items	HKCC 2010/11 (Entry)		HKCC 2010/11 (Exit)		Result from PolyU Pilot (Early 2006)	
		Cronbach alpha	Mean Inter-item correlation	Cronbach alpha	Mean Inter-item correlation	Cronbach alpha	Mean Inter-item correlation
Overall	56	0.9667	0.3417	0.9596	0.2978	0.95	0.25
Creative Thinking	4	0.8025	0.5040	0.7520	0.4312		
Problem Solving	4	0.8207	0.5337	0.7736	0.4607		
Critical Thinking	4	0.7775	0.4662	0.7136	0.3838		
Life Long Learning	4	0.7738	0.4609	0.7345	0.4088		
Leadership	4	0.8305	0.5506	0.7942	0.4911		
Teamwork	4	0.7345	0.4089	0.7072	0.3765		
Communication	4	0.7819	0.4726	0.7792	0.4687	Min: 0.61 Max: 0.91	Min: 0.30 Max: 0.70
EQ & Psy. Wellness	4	0.8012	0.5019	0.7700	0.4556		
Interpersonal Effectiveness	4	0.8096	0.5153	0.7664	0.4506		
Cultural Appreciation	4	0.8715	0.6291	0.8794	0.6458		
Global Outlook	4	0.8216	0.5351	0.7994	0.4991		
Social & National Responsibility	4	0.6908	0.3584	0.5738	0.2518		
Entrepreneurship	4	0.8506	0.5874	0.8366	0.5614		
Health Responsibility	4	0.7355	0.4101	0.7069	0.3762		

Table 3. Reliability of SAARD questionnaire 2011/12 cohort.

Scale	No. of Items	HKCC 2011/12 (Entry)		HKCC 2011/12 (Exit)		Result from PolyU Pilot (Early 2006)	
		Cronbach alpha	Mean Inter-item correlation	Cronbach alpha	Mean Inter-item correlation	Cronbach alpha	Mean Inter-item correlation
Overall	56	0.9636	0.3210	0.9636	0.3210	0.95	0.25
Creative Thinking	4	0.7564	0.4370	0.7564	0.4370		
Problem Solving	4	0.7870	0.4801	0.7870	0.4801		
Critical Thinking	4	0.7117	0.3816	0.7117	0.3816		
Life Long Learning	4	0.7544	0.4344	0.7544	0.4344		
Leadership	4	0.8300	0.5496	0.8300	0.5496		
Teamwork	4	0.7489	0.4272	0.7489	0.4272		
Communication	4	0.7698	0.4553	0.7698	0.4553	Min: 0.61 Max: 0.91	Min: 0.30 Max: 0.70
EQ & Psy. Wellness	4	0.7958	0.4934	0.7958	0.4934		
Interpersonal Effectiveness	4	0.8011	0.5017	0.8011	0.5017		
Cultural Appreciation	4	0.8761	0.6387	0.8761	0.6387		
Global Outlook	4	0.8132	0.5212	0.8132	0.5212		
Social & National Responsibility	4	0.5865	0.2618	0.5865	0.2618		
Entrepreneurship	4	0.8199	0.5323	0.8199	0.5323		
Health Responsibility	4	0.7337	0.4078	0.7337	0.4078		

1) I weigh carefully ethical/moral issues before taking actions.

2) I avoid talking on my mobile phone inside libraries or cinemas.

3) I help those in need by taking part in voluntary social services, money donations, or other activities.

4) I care about what I can do for the good of my country.

In order to have a better understanding of the problem, we calculated at the correlations of the four items of 2011/12 cohort at "Entry" in **Table 4**.

From **Table 4**, we can see that the 2nd item has the least correlation with other items in the scale, indicating that the 2nd item is not suitable for this scale. One possibility for such problem is the extensive scope of the scale Social and National Responsibility. The 2nd item is microscopic in nature as compared to the other items in the scale, which may explain the disparity observed. It is recommended to either drop the 2nd item or revise the item to better align with other items in the same scale.

6. Anomalies in SAARD Survey

To check if there is any anomaly with the responses collected from the SAARD survey, the variances of the 56 questions are plotted sequentially in **Figure 1** for the two cohorts of students at entry and exit (a total of 4 samples). The plot reveals several interesting patterns. First of all, the variations of the responses are strikingly similar for the four samples under study. Consistency across different samples is desirable as we want a survey instrument that produces predictable results over time. However, the plot also reveals some undesirable patterns. We can observe that the variances of the 56 questions are increasing as we progress to the end of the questionnaire. Furthermore, there are several questions have exceptionally large variances, namely question 46, 48 and 55.

The increasing trend in variances can be confirmed by regressing the question variance on question id. As the question variances are computed on the same set of data, a first order auto regressive correlation structure AR(1) is used to adjust for potential correlations across questions. The slope of the model can be interpreted as the rate

of increase in variations of the questions as the respondent proceed through the questionnaire. Such model is fitted to the two cohorts at both Entry and Exit. The slopes of these models are reported in **Table 5**. The slopes in the four models are all highly significant even after the exclusion of questions with extreme variations near the end of the questionnaire. Although the slopes seem small, the change in variance can be significant if the questionnaire is long enough. In our case, the questionnaire has 56 questions, which implies the variance of the last question can be greater that of question one by as much as 0.56 for a slope of 0.01.

The two undesirable features hinted at two potential problems in the survey. The first one is the fatigue effect of working on a lengthy survey. Fatigue is a well-known issue in survey design and it was studied by many survey researchers such as Herzog and Bachman [11]. The second anomaly is the sharp increase in variances of question 46, 48 and 55. To get a perspective of the magnitude of the variations observed, we can compare them to the situation where respondents answering a 7-point Likekert scale question randomly. If responses are distributed randomly over the 7 choices, a simple calculation will yield a theoretical variance of 4. For example, the variance of question 46 is beyond 3, which is getting close to the situation of random responses. The three items with abnormal variations are:

Table 4. Correlations of social and national responsibilities items.

Pearson Correlations	Item 1	Item 2	Item 3	Item 4
Item 1	1.000			
Item 2	**0.2385**	1.0000		
Item 3	0.2873	**0.1417**	1.0000	
Item 4	0.4111	**0.2120**	0.3606	1.0000

Table 5. Rate of increase in question variances.

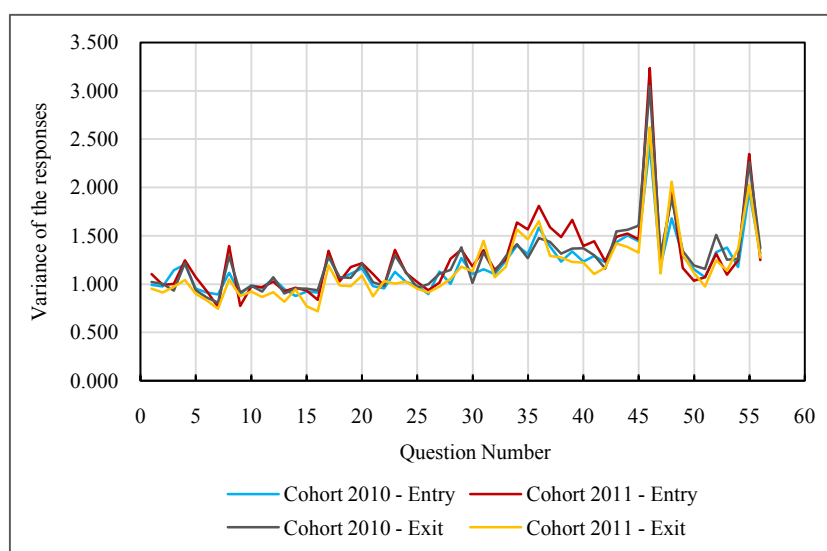| Cohort	Entry/Exit	Slope (p-value)	Slope (p-value) Excluding question 46, 48, 55
2010	Entry	0.010 (<0.001)	0.007 (0.002)
	Exit	0.013 (<0.001)	0.009 (<0.001)
2011	Entry	0.012 (<0.001)	0.007 (0.044)
	Exit	0.013 (<0.001)	0.009 (0.003)

Figure 1. Variances of all questions in different samples.

Item 46: I exercise for 20 - 30 minutes at least twice a week.

Item 48: I engage in short walks or other mild physical activities every week.

Item 55: I avoid talking on my mobile phone inside libraries or cinema.

The large variations to these three items may due to imprecise wording of the questions. For item 46 and 48, both of them involve time duration and quantity of activities under consideration. These can be considered as a double barrel question, which are difficult for respondents to answer. Item 55 may also suffer from a similar problem as the question is asking for the respondent's activities in two different situations, namely libraries and cinema. There is a possibility that the respondents' behaviors in libraries and cinema are very different due to the nature of these environments. A rewording of these questions will therefore be needed. Alternatively, we can consider dropping these 3 questions from SAARD.

7. Construct Validity of SAARD

In this section, we evaluate if the SAARD surveys conducted in HKCC are measuring the same set of constructs. The evaluation is done by comparing the latent factors recovered from HKCC SAARD surveys to that of identified in PolyU pilot study.

In the PolyU pilot conducted in 2006, they identified an 11-factor solution for the SAARD survey. The factors identified together with their corresponding items are listed in **Table 6**.

To facilitate the comparison with PolyU pilot study, exploratory factor analysis will first be applied to the HKCC SAARD survey data. The results are summarized in **Table 7**. For the two HKCC cohorts, percentages of variances explained by the top 11 factors range from 58.46% to 62.21%. On the contrary, the PolyU pilot can only explain a much lower percentage of the variances, 48.54%. This suggests that that an 11-factor model could be sufficient for data collected by HKCC SAARD survey. Alternatively, we also look for important factors with eigenvalues greater than 1. Number of factors identified based on eigenvalue method ranges from 8 to 11. In particular, number of factors identified for 2011 cohort is exactly the same as the PolyU pilot.

Table 6. Factors identified in 2006 PolyU pilot study.

Factors	Items in SAARD
F1 Communication	20, 21, 23, 28
F2 Cultural Appreciation	37, 39, 43, 44
F3 EQ and Psychological Wellness	32, 33, 34, 36
F4 Entrepreneurship	47, 49, 50, 51
F5 Global Outlook	38, 40, 41, 42
F6 Interpersonal Effectiveness	29, 30, 31, 35
F7 Social and National Responsibility	45, 52, 53, 55
F8 Cognitive and Learning Competence	
• Critical Thinking	1, 2, 6, 9
• Creative Thinking	4, 5, 10, 12
• Problem Solving	7, 13, 14, 16
• Life-long Learning	3, 8, 11, 15
F9 Working in Groups	
• Leadership	18, 19, 24, 26
• Teamwork	20, 21, 23, 28
F10 Healthy Lifestyle	
• Engagement in Physical Activity	46, 48
F11 Healthy Lifestyle	
• Health Responsibility	54, 56

Table 7. Exploratory factor analysis results.

Cohort	Entry/Exit	% of variances explained with 11 factors	No. of factors with eigenvalues > 1
2006	N/A	48.54%	Not available
2010	Entry	58.46%	8
	Exit	57.74%	10
2011	Entry	58.48%	11
	Exit	62.21%	11

Table 8. Confirmatory factor analysis results.

Cohort	Entry/Exit	CFI	RMSEA
2006	N/A	0.90	0.040
2010	Entry	0.93	0.040
	Exit	0.90	0.043
2011	Entry	0.87	0.046
	Exit	0.87	0.050

To further validate if a 11-factor model is appropriate, a confirmatory factory analysis is used to assess the goodness of fit of the 11-factor model by using comparative fit index (CFI) and root mean square error of approximation (RMSEA). RMSEA in **Table 8** are all less than or equal to 0.05, which can be considered as good fits according to MacCallumetal [12]. Comparing to the RMSEA of PolyU pilot, the 11-factor model achieved similar fit to the SAARD data collected in HKCC. In terms of CFI, the fits are also comparable to that of PolyU pilot.

8. Conclusions and Recommendations

In this study, we evaluated the reliability and construct validity of the SAARD when applied to the population of associate degree students in HKCC. The results indicated that an 11-factor model can be generalized to the associate degree population in HKCC. However, we also uncovered several deficiencies in the SAARD instrument.

The first issue identified is the potential fatigue effect as respondents proceed through the questionnaire. The effect has a direct impact on the reliability of items near the end of the questionnaire. With the inflated variances of the affected items, the power of hypothesis tests carried out on items near the end of the survey will be lowered. The problem can be alleviated by reducing the length of the survey. A common recommendation in structural equation modeling is to have a minimum of 3 items per construct [13]. Since the SAARD questionnaire is using four items per construct, there is still room for reducing the number of items per construct to three. However, such reduction should not be done by subjective evaluation of the questions alone. One can take into account of the results obtained from the confirmatory factor analysis model. A possibility is to take out the items with the smallest factor loading, in order words, the items with least contribution to their corresponding constructs. In addition to the removal of items least associated with their intended constructs, the questions corresponding to the three erratic items should be revised to alleviate the ambiguities in those questions.

References

[1] Braskamp, L.A., Trautvetter, L.C. and Ward, K. (2006) Putting Students First: How Colleges Develop Students Purposefully. Anker Publishing Company, Bolton.

[2] Braxton, J.M. (2009) Understanding the Development of the Whole Person. *Journal of College Student Development*, **50**, 573-575. http://dx.doi.org/10.1353/csd.0.0100

[3]   Evans, N.J., Forney, D.S., Guido, F.M., Patton, L.D. and Renn, K.A. (2010) Student Development in College. 2nd Edition, Jossey-Bass, San Francisco.

[4]   Kuh, G.D. (2009) What Student Affairs Professionals Need to Know about Student Engagement. *Journal of College Student Development*, **50**, 683-706. http://dx.doi.org/10.1353/csd.0.0099

[5]   Baxter Magolda, M.B. (2009) The Activity of Meaning Making: A Holistic Perspective on College Student Development. *Journal of College Student Development*, **50**, 621-639. http://dx.doi.org/10.1353/csd.0.0106

[6]   Pizzolato, J.E., Brown, E.L., Hicklen, S.T. and Chaudhari, P. (2009) Student Development, Student Learning: Examining the Relation between Epistemologic Development and Learning. *Journal of College Student Development*, **50**, 475-490. http://dx.doi.org/10.1353/csd.0.0093

[7]   Taylor, K.B. (2008) Mapping the Intricacies of Young Adults' Developmental Journey from Socially Prescribed to Internally Defined Identities, Relationships, and Beliefs. *Journal of College Student Development*, **49**, 215-234. http://dx.doi.org/10.1353/csd.0.0005

[8]   Fung, D., Lee, W. and Wong, S.L.P. (2006) Project on Assessing the Development of Generic Competencies of PolyU Students—Report of Findings. Student Affairs Office and Educational Development Centre, Hong Kong Polytechnic University, Hong Kong.

[9]   Fung, D., Lee, W. and Wong, S.L.P. (2007) A New Measure of Generic Competencies. *Key Competencies—Skills for Life* 2007 *Conference*, London, 12-13 July 2007.

[10] So, C.H.J., Lai, S.F.H., Lam, D. and So, Y.L. (2011) Mapping the Impact on Holistic Development: A Study of the Relationship between Generic Skills and Academic Discipline Among Hong Kong Associate Degree Students. *ICERI* 2011 *Proceedings*, Madrid, 14-16 November, 2011, 6630-6638.

[11] Herzog, A.R. and Bachman, J.G. (1981) Effects of Questionnaire Length on Response Quality. *Public Opinion Quarterly*, **45**, 549-559. http://dx.doi.org/10.1086/268687

[12] MacCallum, R.C., Browne, M.W. and Sugawara, H.M. (1996) Power Analysis and Determination of Sample Size for Covariance Structure Modeling. *Psychological Methods*, **1**, 130-149. http://dx.doi.org/10.1037/1082-989X.1.2.130

[13] Hair, J.F., Black, W.C., Babin, B.J. and Anderson, R.E. (2010) Multivariate Data Analysis. 7th Edition, Pearson, New York.

Open Access Library