Global Competency Assessment Scale for Undergraduates in the Contemporary China's Higher Education

Jian Li, Jinhui Xu

Indiana University Bloomington, Bloomington, U.S.A

The purpose of this study is to design global competitive competency assessment scale for undergraduates in China. Cultivating global competency is essential to enhance cross-cultural competitiveness and comprehensive competitive power. Specifically, the global competency is an assessment tool that measures the level of intercultural competency/sensitivity for the individual level for perceiving global cultural capacity in global culture context. By applying item response theory to examine different items contributing cultivating global competency, criterion validity of this assessment was assessed. The results indicate that the global competency assessment scale has strong predictive validity toward the bottom-line goals at an individual level.

Keywords: Global Competency Assessment Scale, contemporary China's higher education, undergraduates, global competency

Introduction

In order to develop into vibrant, mutually beneficial partnerships among nations, gaining an in-depth understanding of cross culture is essential in contemporary world (Merryfield, 2001). Moreover, building effective and positive relationships among different cultures, breaking down the barriers of prejudice, and racism, offering international collaborations are both significant to individual and institutional global competitiveness. Expanding the intercultural competent involves in constructing the intercultural-awareness (Hammer, 2012). The American Association of Colleges and Universities (2007) argued that there are a number of core learning outcomes that is necessary to assist students to face with the challenges of a global society. The development of individual and social responsibility is importance, particularly as it connected to the intercultural knowledge, competence, and engagement. Moreover, Chickering and Braskamp (2009) also highlighted that the conceptual ideas of developing and internalizing a global perspective into individual thinking and identity (p. 27). In order to prepare students more adequately for the challenges of an increasingly diverse and global society, the educators and administrators recognized the importance of global insight or perspective (Braskamp, 2011; Engberg & Hurtado, in press; Hurtado, 2003). In accordance with enhance cross-cultural competitiveness and comprehensive competitive power, it is essential to cultivate global competency. Specifically, this study aims at designing global competitive competency assessment scale for undergraduates in China. This research also focused on generating an assessment tool that measures the level of intercultural competency/sensitivity for the individual level for perceiving global cultural capacity in global culture context.

Literature Review

Generally speaking, global competence can improve organizational and individual effectiveness in learning and career success. The importance of international communication knowledge and skills is considered to guarantee career success in global context. In the globally integrated labor market, striving to obtain global competence can ensure the organizational and individual competitiveness and effectiveness. It seemed plausible to encourage students to utilize global competence to cope with different issues in response to different foreign tasks.

The interconnection among politics, culture, environment and technologies witnesses dramatic increases during recent decades (Merryfield, 2008). The interconnection is associated with the international trades, migration flow, and effective communications. Complex global issues also influence the development of regional economic conflicts. So, mastering a multidisciplinary and global knowledge can be considered as one crucial factor to competitive with others (Lohmann & Rollins, 2006). However, according to recent studies of American Council on Education (ACE), they found that few higher education institutions emphasized on improving internationalization and globalization levels. In addition, National Research Council also indicated that improving student global competency is important to cultivate global corporate skills (Reimers, 2009). The cross-cultural and globally oriented education consistently encourages higher education institutions to cultivate global citizens with intercultural sensitivity and global competency. Cross-cultural study can always be regarded as one main approach to deepen students' global perspectives and competences (Hill, 1991). Alternative pedagogical management should cultivate student global competence in the long term. In higher education system, the student global competence is always considered as significant learning outcome. However, there is still much more controversial to define the ideas of global competence and perspective. Additionally, how to design and implement pedagogical interventions to promote student global competency are still confused by policymakers in current global education background (Fantini, 2009).

There existed little effort to standardize the evaluation or assessment of global education in US. Godbey (2002) indicated that universities graduates were lack of global knowledge and skills to enable them to be global graduates. Deardorff (2004) also argued that few American institutions enable cultivate global undergraduates. Green (2000) also made similar conclusion that few American college graduates are fully competent to perform in different kinds of cultures, speak foreign language, or holding some significant understanding of the world worldwide. Gliozzo (2002) also noted that Americans were lack of the most global knowledge and skills, which are necessary to act in multicultural environments. In order to qualify the global competent citizens, the ACE suggested that America's future should concentrate on developing a global citizen by training advance foreign language and intercultural knowledge. ACE (2000) further indicated that mastery of intercultural competency and global knowledge is an important component of an effective international education. Curran (2003) suggested that the global competent is the capacity to be familiar with the global environment. Moreover, Wilson and Dalton (1997) concluded that the perceptual knowledge (open-mindedness, resistance to stereotyping, complexity of thinking, and perspective consciousness) and substantive knowledge (of cultures, languages, world issues, global dynamics, and human choices) were integral parts of delimiting global competence. The higher education should has indispensable responsibility to educate students to be knowledgeable and responsible as they are engaged in the international context to learn foreign languages, to be knowledgeable of the across-culture, the economics and policies of the countries they will visit, to interact in a intercultural way (Rimer, 2004). In accordance with seeking to expand the intercultural horizons and improving

intercultural opportunities, most of countries, such as UK, France, Italy, and Australia, implemented effective relevant policy to stimulate the development of global education.

The intercultural/across cultural is more and more becoming a institutional recognition to design global curricular program and projects in terms of improving the global competency institutionally. There exists less research that defining the concept of global competency in response to the rapid increasing number of international student mobility across China. Within the limitation of this research, there is no relevant consensus for undergraduates in China. In order to provide effective suggestion in improving global competency for China's institutions, this research has been undertaken to development the new definitions of competitive global competency for the contemporary China's higher education system. The ultimate purpose of this research is to encourage institutional administrators to design and create a global competitive curriculum and provide effective and sufficient actives to guarantee individual competent globally.

Methodology

Descriptive Data Analysis

This survey aims at the study of Chinese undergraduate students from five grades, thirteen majors in eight universities from Beijing, China. The total sample size is 2,505. All the respondents answered this questionnaire by paper-based survey. In this study, we applied SPSS 23.0, Stata 14.1 and Conquest 4.0 to analyze the variables. Specifically, Global Competency Assessment Questionnaire with Rating Scale has been entitled in this survey. In order to guarantee the reliability and validity of this rating scale questionnaire, we designed the questionnaire concerning on the definition of the dimensions we adopt. We used 5-point Likert scale to score the answers of each question. The respondents answered all the questions in the questionnaire relating to personal psychical cognition. In this study, the rating scale questionnaire includes five sections, which included personal social background, global knowledge, global skill, global attitude and global experience. Moreover, comparing with the traditional survey and questionnaire concerning on student's global competency, we add global experience as our new dimension in consistent with our theoretical framework (see Figure 1).

Figure 1. Global competency assessment scale.

This survey contains 13 questions concerning on personal social and educational background information, which includes gender, university, discipline (major), year of the grade, university classification, family location, whether respondent had intercultural training or internship, whether has global experience, parents' education background, family monthly income, demographic and academic performance (see Table 1).

Table 1

Descriptive Data of the Study

Variable	<i>N</i>	Percentage (%)
Sex	2,505	
Male	1,260	50.30
Female	1,245	49.70
University	2,505	
Beihang University	852	34.01
Capital Normal University	297	11.86
Beijing Normal University	292	11.66
Beijing City University	278	11.10
China Agricultural University	272	10.86
Beijing University of Technology	242	9.66
China University of Geosciences	218	8.70
China Institute of Industrial Relations	52	2.16
Discipline	2,505	
Engineering	1,323	52.81
Science	277	11.06
Education	273	10.90
Economics	182	7.27
Management	165	6.59
Law	62	2.48
Literature	57	2.28
Agriculture	31	1.24
Philosophy	15	0.60
History	6	0.24
Arts	5	0.20
Medicine	3	0.12
Else	105	4.19
Grade	2,503	
Freshmen	545	21.77
Sophomore	846	33.80
Junior	701	28.01
Senior	409	16.34
Fifth-year Student	2	0.08
University classification	2,505	
985-Project University	1,416	56.53
211-Project University	491	19.60
Others	598	23.97
Family location	2,498	
Urban	1,774	71.02
Rural	724	28.98
Intercultural training or internship	2,498	
Have no experience for intercultural training or internship	2,045	81.87
Have experience for intercultural training or internship	453	18.13
Global experience	2,495	
Have no experience for intercultural training or internship	2,169	86.93
Have experience for intercultural training or internship	326	13.07
Father's education background	2,489	
Not be educated	21	0.84
Elementary School	156	6.27
Middle School	512	20.57
High School or Secondary Specialized School	803	32.26

Table 1 continued

Variable	<i>N</i>	Percentage (%)
University/College	877	35.24
Master	84	3.37
Ph.D	36	1.45
Mother's education background	2,488	
Not be educated	54	2.17
Elementary School	259	10.41
Middle School	536	21.54
High School or Secondary Specialized School	776	31.19
University / College	779	31.31
Master	60	2.41
Ph.D	24	0.96
Family monthly total income (RMB, unit: Yuan)	2,398	
Below 1,000	146	6.09
1,000-2,999	388	16.18
3,000-5,999	799	33.32
6,000-9,999	617	25.73
10,000-14,999	239	9.97
15,000-19,999	95	3.96
20,000-29,999	49	2.04
Above 30,000	65	2.71
Demographic	2,273	
Beijing	555	24.42
Shandong	138	6.07
Henan	113	4.97
Hebei	109	4.80
Hubei	84	3.70
Shanxi	80	3.52
Inner Mongolia	78	3.43
Shaanxi	77	3.39
Sichuan	75	3.30
Anhui	74	3.26
Hunan	74	3.26
Heilongjiang	65	2.86
Zhejiang	65	2.86
Jiangxi	62	2.73
Guizhou	60	2.64
Gansu	59	2.60
Xinjiang	56	2.46
Liaoning	55	2.42
Jilin	53	2.33
Chongqing	48	2.11
Fujian	46	2.02
Tianjin	45	1.98
Jiangsu	44	1.94
Yunnan	34	1.50
Guangxi	30	1.32
Guangdong	24	1.06
Ningxia	23	1.01
Qinghai	15	0.66
Shanghai	14	0.62
Hainan	12	0.53
Tibet	5	0.22
Hong Kong	1	0.04
Academic performance last year	2,491	
Previous 25%	713	28.62
25%-50%	658	26.42
50%-75%	448	17.98
Post 75%	212	8.51
Unknown	460	18.47

Dimensions of Global Competency Assessment Questionnaire with Rating Scale

There are totally 45 questions in this macro section with 4 dimensions—global knowledge, global skill, global attitude and global experience. In another words, all these four dimensions contribute to assess and evaluate the global competency.

Global Knowledge

There are totally 11 questions: item 14 to 24. We use 5-point Likert scale for scoring: (1) The score of strongly disagree is one; (2) The score of disagree is two; (3) The score of unsure is three; (4) The score of agree is four; (5) The score of strongly agree is five.

Global Skill

In this section, we try to know the actual intercultural skills of respondent. There are totally 9 questions: item 25 to 30, 35 to 36 and 49. We use 5-point Likert scale for scoring. For item 25 to 30, 35 to 36, (1) The score of strongly disagree is one; (2) The score of disagree is two; (3) The score of average is three; (4) The score of agree is four; and (5) The score of strongly agree is five. For item 49, (1) The score of very low is one; (2) The score of quite low is two; (3) The score of unsure is three; (4) The score of quite high is four; and (5) The score of very high is five;

Global Attitude

We try to know the respondent's attitude towards inter-culture. There are totally 15 questions: item 31 to 33 and 37 to 48 (including a counter question, item 48). We use 5-point Likert scale for scoring: (1) The score of strongly disagree is one; (2) The score of disagree is two; (3) The score of unsure is three; (4) The score of agree is four; and (5) The score of strongly agree is five.

Global Experience

In this part, we try to know the actual intercultural experiences or trainings of respondent. There are totally 10 questions: item 34 and 50 to 58. We used 5-point Likert scale for scoring. For item 34, (1) The score of strongly disagree is one; (2) The score of disagree is two; (3) The score of unsure is three; (4) The score of agree is four; and (5) The score of strongly agree is five. For item 50 to 56 and 58, (1) The score of zero time is one; (2) The score of once is two; (3) The score of twice is three; (4) The score of three times is four; and (5) The score of four times or more is five. For item 57, (1) The score of none is one; (2) The score of two months or longer is two; (3) The score of once a month is three; (4) The score of twice a month is four; and (5) The score of once a week is five (see Table 2).

Table 2

The Question Summary of the Scales

Dimension	Amount of questions
Knowledge	11
Skill	9
Attitude	15
Experience	10
Total	45

Reliability and Validity Analysis

The reliability and validity of this survey have been reported as follows: the coefficient of the internal reliability for respondents, the whole survey's Cronbach alpha equals to 0.915, the Cronbach alpha for global

knowledge, skill, attitude and experience equal to 0.886, 0.848, 0.869, and 0.754, respectively, which demonstrated the survey data for the respondents has great internal reliability. In addition, analyzing the validity of the data. We obtain that the value of Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy equals to 0.937, and the p -value for Bartlett's test of sphericity is equal to 0, which indicates that the data is of a great validity and suitable within the data, and it is suitable to make the further analysis with both questionnaire and respondents' global competency.

Data Analysis

Item Response Theory

Item response theory (IRT) is considered as a paradigm for designing, analyzing, and scoring of tests, questionnaires, and similar instruments measuring abilities, attitudes, or other variables, which emerged as early as the 1940s though the popularity came much later in the 1970s. It is based on the concept of the probability of a correct/keyed response to an item, which is a mathematical function of person and item parameters. The person parameter is construed as (usually) a single latent trait or dimension. Broadly speaking, IRT models can be divided into two bodies: uni-dimensional and multidimensional.

Specifically, Uni-dimensional IRT Models for Dichotomous Responses hold two fundamental assumptions: uni-dimensionality and local independence. The assumption of uni-dimensionality is related to a set of items and/or a test measure(s) only one latent trait (θ), and local independence refers to the assumption that there is no statistical relationship between examinees' responses to the pairs of items in a test, once the primary trait measured by the test is removed. The two assumptions are really just different approaches and the third main assumption focused on modeling the relationship between the trait measured by the test and item responses. Historically, there are some classical models: the normal ogive model was regarded as the first IRT model focusing on measuring psychological and/or educational latent traits (Richardson, 1936; Mosier, 1940, 1941; Ferguson, 1942; Lawley, 1943). In the model, an item characteristic curve (ICC) is derived from the cumulative density function (CDF) of a normal distribution; One-Parameter Logistic Model (1PLM/Rasch Model) is related to a mathematician in Denmark, George Rasch, followed up with a various approach to IRT in the 1950s (Rasch, 1960(1980); 1961; 1967; 1968). He used a logistic function to derive an ICC instead of the normal ogive function and his model contributed to simplifying the normal ogive model and the complexity of computation; Two-Parameter Logistic Model (2PLM) is a generalization of the 1PLM. Instead of having a fixed discrimination of "1" across all items as in 1PLM, in the 2PLM, each item has its own discrimination parameter; Three-Parameter Logistic Model (3PLM) provides an ICC to have non-zero lower asymptotes and this model is more suitable for response data with those items in which examinees at the extremely low proficiency level may get the items correctly by chance; Nonparametric Item Response Model (ICCs) are characterized by a single function in IRT models with parameters. Additionally, Ramsay (1991) proposed a kernel smoothing approach for nonparametric item response models. Nonparametric item response models may not be as practically useful for operational uses as parametric models because nonparametric item response models do not provide informative, interpretable item parameters, and it is hard to equate tests under nonparametric models. However, nonparametric models are frequently used for research purposes such as evaluating model fit for parametric models since nonparametric models produce item characteristic functions that are very close to the observed data. Moreover, there also existed some specific models generated by many scholars consistently: Partial Credit Model (PCM) is an extension of the 1PLM (a.k.a., Rasch model) (Masters,

1982; 1986; 1988a; 1988b); Generalized Partial Credit Model (GPCM) (Muraki, 1992) is a generalization of the PCM with a parameter for item discrimination added to the model; Graded Response Model (GRM) was introduced by Samejima (1968; 1972; 1995) to handle ordered polytomous categories such as letter grading, A, B, C, D, and F and polytomous responses to attitudinal statements (such as a Likert scale); Nominal Response Model (NRM) was also introduced by Bock (1972). Therefore, the values of the responses do not represent some sort of scores on items, but just nominal indications for response categories. Some applications of the NRM are found in uses with multiple choice items. Additionally, Multidimensional Compensatory Three-Parameter Logistic Model (MC3PLM) can be seen as an extension of the uni-dimensional 3PL model. Moreover, IRT is suitable for Likert-scale tradition (Andrich, 1978).

Rating Scale Model (RSM)

Rating Scale Model (RSM) is one of the models of Rasch family, always applied in the rating scale survey. There are two different approaches to the rating scale model. Andersen's (1977; 1983) proposed a response function, in which the values of the category scores are directly used as a part of the function:

$$P_{ix}(\theta) = \frac{e^{w_i\theta - a_{ix}}}{\sum_{x=1}^m e^{w_x\theta - a_{ix}}}$$

where $w_1, w_2, \dots, w_m$ are the category scores, which prescribe how the m response categories are scored, and a_{ih} are item parameters connected with the items and categories. An important assumption of this model is that the category scores are equidistant. Linear Rating Scale Model (LRSM) was proposed to generalized Rasch model and rating scale model (Fischer & Parzer, 1991). Since the whole scale is used 5-point Likert scale, which is also ordered, we adopt the RSM we introduced as our analyzing method.

Fundamental Estimation: The Analysis and Estimation of Item Difficulty

In application, we usually set the difficulty of the item from the range of -3 to 3. The larger the value, represents the more difficult in this item; the smaller the value, represents the easier of the item (Yu, 2009).

Test of Goodness-of-Fit

We used the indexes T -value and MNSQ (Fit mean-square) to judge the goodness of the items. For T -value, with 95% significant level, the range of the value is from -1.96 to 1.96. Out of this range means the answers of the item has the significant difference between what we expect and reality. MNSQ is the standardized mean-square residual summary statistics, describing the variance between observation values and expected values for all respondents' answers of the item. For rating scale model, the range of the MNSQ value is from 0.6 to 1.4, which means the respondent's response conforms to the what we expect of the item. Informal simulations studies and experience analyzing hundreds of datasets indicate that when MNSQ is larger than 2, it distorts or degrades the measurement system; when MNSQ is from 1.4 to 2, it is unproductive for construction of measurement, but not degrading; when MNSQ is smaller than 0.6, it is less productive for measurement, but not degrading, which may produce misleadingly good reliabilities and separations. Totally, when MNSQ is larger than 1.4, it indicates that the item measures the other specialties rather than the dimension now; when MNSQ is smaller than 0.6, it indicates that the item may be covered by the other items. Specifically, when MNSQ equals to 1, it means the response completely conforms to expectation (Wright, Linacre, Gustafson, & Martin-Lof, 1994).

The Analysis and Estimation of Category Difficulty (Strata)

The value of the estimation of category difficulty ranges from low to high, the higher value of the estimation means the more difficult that the respondent reaches the standard (Yu, 2009). Since in our study, we

use 5-point Likert scale, we rationally set four categories in this case. The larger the estimation coefficient, the harder that the respondent reaches this category; the smaller the estimation coefficient, the easier the respondent reached this category. The category also represents the ability level of the total ability.

Scale Reliability Analysis

In our study, we adopt EAP/PV reliability index to test the reliability of our rating scale. EAP/PV reliability is explained variance according to the estimated model divided by total person variance (Adams, 2006) (Wu, Adams, & Wilson, 2007). The EAP/PV reliability estimates reported comes from an analysis of all student responses from across the forms, which includes all of the item data missing due to the test design, and thus gives an underestimate of the reliability for those constructs that one would expect in a normal administration (Schwartz, 2012). Also, EAP/PV reliability can be interpreted like Cronbach's alpha (Draney & Wilson, 2008). However, It is based on the item response model estimates rather than the raw score, and it is useful in situations where there is too much planned missing data for Cronbach's alpha to be estimated (Scalise, Madhyastha, Minstrell, & Wilson, 2010).

RSM Analysis

Firstly, we used Conquest 4.0 to analyze out scale data. The total output and the outputs for each dimension have been shown from Table 3 to Table 7, the item separation reliability (ISR) is 0.998. ISR gives the test user an indication of how well items are separated by the persons taking the test. It equals to

$$ISR = 1 - \left[\frac{MSE_i}{SD_i^2} \right]$$

where MSE_i is the mean square error of the item and SD_i is the standard deviation of the item. In this case, it means all the items in the survey are sufficiently well separated in difficulty to identify the direction and meaning of the variable.

Table 3

The Estimated Coefficients for Items and Total Numbers of Respondents for Each Option in the Questionnaire

Item	Estimate	Error	MNSQ	T	Respondent	Respondent for each option				
						1	2	3	4	5
14	0.351	0.017	0.90	(-3.5)	2,500	180	650	780	743	147
15	0.246	0.017	0.74	(-10.1)	2,503	118	592	867	809	117
16	-0.080	0.017	0.74	(-10.1)	2,500	103	472	656	1,078	191
17	-0.181	0.017	0.84	(-6.1)	2,502	79	407	699	1,115	202
18	0.457	0.017	0.77	(-8.9)	2,502	136	703	945	627	91
19	0.069	0.017	0.75	(-9.7)	2,502	97	510	815	930	150
20	0.004	0.017	0.84	(-5.8)	2,501	107	471	757	1,005	161
21	-0.269	0.018	0.78	(-8.2)	2,503	90	297	753	1,149	214
22	-0.250	0.018	0.74	(-10.1)	2,501	66	308	813	1,119	195
23	-0.075	0.017	0.78	(-8.3)	2,499	75	376	903	987	158
24	-0.269*	0.055	1.07	(2.4)	2,502	83	382	663	1,100	274
25	-0.365	0.017	0.87	(-5.0)	2,501	151	532	862	790	166
26	-0.122	0.016	1.13	(4.3)	2,500	240	682	739	673	166
27	-0.045	0.016	0.78	(-8.4)	2,502	202	746	834	590	130
28	0.155	0.016	0.72	(-11.1)	2,501	234	840	875	471	81
29	0.314	0.016	0.74	(-10.0)	2,503	310	908	823	369	93
30	-0.007	0.016	0.88	(-4.6)	2,499	218	729	862	586	104

Table 3 continued

Item	Estimate	Error	MNSQ	<i>T</i>	Respondent	Respondent for each option				
						1	2	3	4	5
31	0.729	0.017	0.99	-0.3	2,496	144	574	728	921	129
32	0.324	0.017	0.90	(-3.5)	2,497	95	371	663	1,181	187
33	1.047	0.016	1.03	1.1	2,501	186	666	1,016	535	98
34	-0.774	0.019	1.58	(17.4)	2,499	589	978	471	383	78
35	-0.001	0.016	0.77	(-8.8)	2,501	227	637	1,009	548	80
36	-0.019	0.016	0.80	(-7.6)	2,502	205	617	1,088	507	85
37	0.126	0.017	0.86	(-5.2)	2,504	82	298	667	1,177	280
38	-0.260	0.017	0.82	(-6.9)	2,501	72	201	460	1,345	423
39	-0.130	0.017	0.86	(-5.1)	2,503	85	222	567	1,200	429
40	-0.057	0.017	0.89	(-4.2)	2,503	86	242	605	1,154	416
41	-0.237	0.018	0.69	(-12.4)	2,501	66	173	508	1,346	408
42	-0.312	0.018	0.70	(-11.9)	2,504	52	154	456	1,433	409
43	-0.433	0.018	0.72	(-11.2)	2,504	51	126	392	1,455	480
44	-0.404	0.018	0.79	(-8.2)	2,502	49	136	402	1,427	488
45	-0.682	0.018	1.10	(3.3)	2,503	41	107	354	1,274	727
46	-0.463	0.018	1.05	1.9	2,503	44	112	470	1,301	576
47	-0.386	0.018	1.00	0.0	2,501	51	114	485	1,355	496
48	1.138*	0.065	3.30	(51.9)	2,503	664	482	341	613	403
49	0.089*	0.047	0.79	(-8.2)	2,494	219	491	1,447	295	42
50	-0.552	0.019	1.41	(13.0)	2,494	770	849	580	168	127
51	0.201	0.021	1.40	(12.5)	2,494	1,336	580	417	98	63
52	0.179	0.021	1.76	(21.9)	2,494	1,324	660	314	86	110
53	0.369	0.021	1.36	(11.5)	2,493	1,422	627	284	79	81
54	-0.249	0.020	1.63	(18.7)	2,492	1,126	733	289	125	219
55	0.108	0.021	1.42	(13.0)	2,429	1,306	546	352	121	104
56	1.379	0.024	1.37	(11.5)	2,429	1,919	291	152	45	22
57	-1.078	0.019	3.27	(50.7)	2,430	1,278	136	135	148	733
58	0.417*	0.062	1.40	(12.5)	2,430	1,393	688	180	61	108

Notes. * It is constrained; “_” It marks the value of MNSQ exceeds the range of (0.6, 1.4); “()” It marks the *T*-value that exceeds the range of (-1.96, 1.96).

Table 4

The Estimated Coefficients for Knowledge and Total Numbers of Respondents for Each Option

Item	Estimate	Error	MNSQ	<i>T</i>	Respondent	Respondent for each option				
						1	2	3	4	5
14	0.351	0.017	0.90	(-3.5)	2,500	180	650	780	743	147
15	0.246	0.017	0.74	(-10.1)	2,503	118	592	867	809	117
16	-0.080	0.017	0.74	(-10.1)	2,500	103	472	656	1,078	191
17	-0.181	0.017	0.84	(-6.1)	2,502	79	407	699	1,115	202
18	0.457	0.017	0.77	(-8.9)	2,502	136	703	945	627	91
19	0.069	0.017	0.75	(-9.7)	2,502	97	510	815	930	150
20	0.004	0.017	0.84	(-5.8)	2,501	107	471	757	1,005	161
21	-0.269	0.018	0.78	(-8.2)	2,503	90	297	753	1,149	214
22	-0.250	0.018	0.74	(-10.1)	2,501	66	308	813	1,119	195
23	-0.075	0.017	0.78	(-8.3)	2,499	75	376	903	987	158
24	-0.269*	0.055	1.07	(2.4)	2,502	83	382	663	1,100	274

Notes. * It is constrained; “()” It marks the *T*-value that exceeds the range of (-1.96, 1.96).

Table 5

The Estimated Coefficients for Skill and Total Numbers of Respondents for Each Option

Item	Estimate	Error	MNSQ	T	Respondent	Respondent for each option				
						1	2	3	4	5
25	-0.365	0.017	0.87	(-5.0)	2,501	151	532	862	790	166
26	-0.122	0.016	1.13	(4.3)	2,500	240	682	739	673	166
27	-0.045	0.016	0.78	(-8.4)	2,502	202	746	834	590	130
28	0.155	0.016	0.72	(-11.1)	2,501	234	840	875	471	81
29	0.314	0.016	0.74	(-10.0)	2,503	310	908	823	369	93
30	-0.007	0.016	0.88	(-4.6)	2,499	218	729	862	586	104
35	-0.001	0.016	0.77	(-8.8)	2,501	227	637	1,009	548	80
36	-0.019	0.016	0.80	(-7.6)	2,502	205	617	1,088	507	85
49	0.089*	0.047	0.79	(-8.2)	2,494	219	491	1,447	295	42

Notes. * It is constrained; “()” It marks the T-value that exceeds the range of (-1.96, 1.96).

Table 6

The Estimated Coefficients for Attitude and Total Numbers of Respondents for Each Option.

Item	Estimate	Error	MNSQ	T	Respondent	Respondent for each option				
						1	2	3	4	5
31	0.729	0.017	0.99	-0.3	2,496	144	574	728	921	129
32	0.324	0.017	0.90	(-3.5)	2,497	95	371	663	1,181	187
33	1.047	0.016	1.03	1.1	2,501	186	666	1016	535	98
37	0.126	0.017	0.86	(-5.2)	2,504	82	298	667	1,177	280
38	-0.260	0.017	0.82	(-6.9)	2,501	72	201	460	1,345	423
39	-0.130	0.017	0.86	(-5.1)	2,503	85	222	567	1,200	429
40	-0.057	0.017	0.89	(-4.2)	2,503	86	242	605	1,154	416
41	-0.237	0.018	0.69	(-12.4)	2,501	66	173	508	1,346	408
42	-0.312	0.018	0.70	(-11.9)	2,504	52	154	456	1,433	409
43	-0.433	0.018	0.72	(-11.2)	2,504	51	126	392	1,455	480
44	-0.404	0.018	0.79	(-8.2)	2,502	49	136	402	1,427	488
45	-0.682	0.018	1.10	(3.3)	2,503	41	107	354	1,274	727
46	-0.463	0.018	1.05	1.9	2,503	44	112	470	1,301	576
47	-0.386	0.018	1.00	0.0	2,501	51	114	485	1,355	496
48	1.138*	0.065	3.30	(51.9)	2,503	664	482	341	613	403

Notes. * It is constrained; “—” It marks the value of MNSQ exceeds the range of (0.6, 1.4); “()” It marks the T-value that exceeds the range of (-1.96, 1.96).

Table 7

The Estimated Coefficients for Experience and Total Numbers of Respondents for Each Option.

Item	Estimate	Error	MNSQ	T	Respondent	Respondent for each option				
						1	2	3	4	5
34	-0.774	0.019	1.58	(17.4)	2,499	589	978	471	383	78
50	-0.552	0.019	1.41	(13.0)	2,494	770	849	580	168	127
51	0.201	0.021	1.40	(12.5)	2,494	1,336	580	417	98	63
52	0.179	0.021	1.76	(21.9)	2,494	1,324	660	314	86	110
53	0.369	0.021	1.36	(11.5)	2,493	1,422	627	284	79	81
54	-0.249	0.020	1.63	(18.7)	2,492	1,126	733	289	125	219
55	0.108	0.021	1.42	(13.0)	2,429	1,306	546	352	121	104
56	1.379	0.024	1.37	(11.5)	2,429	1,919	291	152	45	22
57	-1.078	0.019	3.27	(50.7)	2,430	1,278	136	135	148	733
58	0.417*	0.062	1.40	(12.5)	2,430	1,393	688	180	61	108

Notes. * It is constrained; “—” It marks the value of MNSQ exceeds the range of (0.6, 1.4); “()” It marks the T-value that exceeds the range of (-1.96, 1.96).

Statistically, we observed that (1) for all items, their difficulty is totally in the range of -3 to 3, which means the total difficulty of the survey is quite suitable to respondents; (2) The MNSQ for global knowledge and global skill dimensions are totally within the range of 0.6-1.4, which indicates the response from the respondent conforms to expectation; for attitude dimension, just item 48 has a very large value of MNSQ, which indicates it tends to measure the other specialties of the respondent. Moreover, item 41 is a very good item since its answer is totally conforming to the expectation with MNSQ equals to 1; for the global experience dimension, the most of the items' MNSQ are larger than 1.4, which indicates that these items tend to measure the other specialties rather than global experience to the respondent; and (3) almost for all the items except some items in the dimension of attitude the *T*-values are not at the range of -1.96 to 1.96, which means that there are significant differences between the actual situation of respondent and expectation.

Estimation of the Coefficients of Each Category's Difficulty

Statistically speaking, we observed that it is easiest for respondent to reach category 1, and the category 4 is the hardest level, which indicates that very few people have the comprehensive ability of global competency. Moreover, the MNSQ for each category are all larger than 1.4 indicates the variance of the responses from respondents are very difference from each category; *T*-value all exceed the range of -1.96 to 1.96, which can indicate that there are significant differences between the actual situation and expectation (see Table 8).

Table 8

The Estimated Coefficients For Category Difficulty

Category	Estimate	Error	MNSQ	<i>T</i>
0			2.79	(43.3)
1	-1.698	0.008	2.73	(42.3)
2	-0.729	0.007	3.34	(52.5)
3	0.132	0.007	3.55	(55.8)
4	2.295*		5.13	(77.0)

Notes. * It is constrained; “—” it marks the value of MNSQ exceeds the range of (0.6, 1.4); “()” It marks the *T*-value that exceeds the range of (-1.96, 1.96).

Distribution Figure of Respondent's Ability and Item Difficulty

Specifically speaking, as shown in Figure 2, we can find that, on the left, the “X” means that the ability distribution for respondents' while on the right it is the difficulty distribution of each item. Moreover, this histogram illustrates that the distribution of the respondents' achievement and items indicate their difficulty level. In this research, each “X” represents 26 respondents. And the respondent's ability distribution for all dimensions is to nearly satisfy the normal distribution. Therefore, we observed that the average ability of respondents' attitude is the largest, close to 1, the most of them are within the range from 0 to 2; the average ability of respondents' knowledge is the second largest, above 0, the most of them are in the range of -1 to 1; the average ability of respondents' skill is the third largest, lightly below 0, the most of them are in the range of -1.5 to 0.5. These three dimensions are very close, which indicates that the ability of these three dimensions of respondents is closed to each other. However, the average ability of respondents' experience is the smallest and has a great distance with the other 3 dimensions, nearly -2. Furthermore, we found that the average difficulty for all the items is around 0 with nearly normal distribution, and most of them are range with -1 to 1, which also indicates that the scientificity of the items design. However, there is still a space among items, which indicates that there is lack of items for the matching ability respondent to answer. Specifically, item 56 is the most

difficult item, which indicates that the respondent tended to select the low score option while 57 is the easiest item, which indicates that the respondent tended to select the high score option.

	Knowledge	Skill	Attitude	Experience	Item
			X		
3			X		
			X		
	X		X		
	X		X		
			XX		
2	X		XXX		
	X		XXX		
	XX	X	XXXX		
	XX	X	XXXXX		
	XXXX	X	XXXXXX		56
	XXXXX	XX	XXXXXXXX		33 48
1	XXXXXXXX	XX	XXXXXXXXXX		
	XXXXXXXX	XXXX	XXXXXXXXXX	X	31
	XXXXXXXX	XXXX	XXXXXXXXXX	X	18
	XXXXXXXX	XXXXXX	XXXXXXXXXX	X	14 29 32 53 58
	XXXXXXXX	XXXXXX	XXXXXX	X	15 19 28 37 49 51 52 55
0	XXXXXXXX	XXXXXXXXXX	XXXXXX	XX	16 20 23 26 27 30 35 36 39 40
	XXXXXXXX	XXXXXXXXXX	XXX	XXX	17 21 22 24 38 41 42 44
	XXXXXX	XXXXXXXXXX	XXX	XX	25 43 44 46 47
	XXXX	XXXXXXXXXX	XX	XXX	45 50
	XXXX	XXXXXXXXXX	X	XXXX	34
-1	XXX	XXXXXXXX	X	XXXX	57
	XX	XXXXX		XXXXXX	
	X	XXXX		XXXXXX	
	X	XX	X	XXXXXX	
	X	XX		XXXXXX	
-2		X		XXXXXX	
		X		XXXXXX	
	X			XXXXXX	
		X		XXXXX	
				XXXXX	
				XXXXX	
-3				XXX	
				XX	
				XXX	
				XX	
				XX	
-4				X	
				X	

Figure 2. The distribution figure for respondents' ability and item difficulty (Each "X" represents 26.0 cases).

Reliability Verification

As shown in Table 9, we observed that the knowledge has the closest relationship with skill among these dimensions with the correlation coefficient equals to 0.634 while the attitude and experience has the smallest correlation coefficient equals to 0.281, which is in accordance with the previous scholar's work. As we can see in Table 10, for each dimension, the EAP/PV reliability is larger than 0.74, indicating that estimates of respondent ability in each dimension have good precision.

Table 9

Correlation Coefficient among Dimensions

Dimension	Knowledge	Skill	Attitude	Experience
Knowledge				
Skill	0.634			
Attitude	0.566	0.656		
Experience	0.350	0.594	0.281	

Table 10

Variance and EAP/PV Reliability of Each Dimension

Dimension	Knowledge	Skill	Attitude	Experience
Variance	0.927	0.814	0.900	1.502
Reliability	0.784	0.779	0.806	0.741
Numbers	11	9	15	10

Results and Findings

Based on the analysis from Rating Scale Model (RSM), for other dimensions, the items have the good fitness to respondents while the items of global experience are not fit to the respondents because the respondents are lacking in global experience while the average difficulty of the global experience items are larger than their ability; The expectation of respondent's global competency is significantly different from the actual situation; Respondents have the high ability in global attitude with the low ability in global experience; The ability of different category's respondents have significantly differences. It is the reason why it is very hard to enter the highest ability category. Fixing some items in the global experience dimension, concentrating on offering some items to fit the ability of respondents can be considered as effective approaches to improving the overall efficiency of Global Competency Assessment Questionnaire with Rating Scale. Additionally, based on the results, we find that the ability of respondents' experience is extremely low and we should fix some items in the global experience dimension. Moreover, since there is still a space in the item difficulty distribution, it is necessary to add some new items, which can fit the ability of respondents. In conclusion, the global competitive competency assessment scale for undergraduates in China is essential to cultivate global competency for graduate students in the contemporary China's higher education. It is also significant to enhance cross-cultural competitiveness and comprehensive competitive power. Specifically, this global competency assessment scale is a suitable assessment tool that measures the level of intercultural competency/sensitivity for the individual level for perceiving global cultural capacity in global culture context. The results indicate that the global competency assessment scale has strong predictive validity toward the bottom-line goals at an individual level.

Limitations and Future Research Suggestions

The key limitation of this research is that the questions of the global experience section are more difficult than the actual ability/level of the respondents'. Hence, generating more effective approach concerning on how to design the questions in this part (Global Experience) is essential to measure the ability of the respondents' comprehensively. Moreover, the suggestion for the future research is to design some new questions of the global experience section that will meet the ability/level of the respondents by applying IRT method to re-test the validity and reliability of the whole scales. Additionally, in order to get in-depth understanding and

evaluating global competency scales, more and more fundamental research should be conducted to enrich the field of measuring different types of global competency for different goals and audience.

References

- Adams, R. J. (2006). Reliability and item response modeling: Myths, observations and applications. In *13th International Objective Measurement Workshop* (pp. 5-7), Berkeley, C.A., April.
- Andersen, E. B. (1977). Sufficient statistics and latent trait models. *Psychometrika*, 42(1), 69-81.
- Andersen, E. B. (1983). Latent trait models. *Journal of Econometrics*, 22(1-2), 215-227.
- Andrich, D. (1978). Scaling attitude items constructed and scored in the Likert tradition. *Educational and Psychological Measurement*, 38(3), 665-680.
- Association of American Colleges and Universities. (2007). *College learning for the new global century: A report from the National Leadership Council for Liberal Education and America's Promise*. Washington, D.C.: Author
- Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37(1), 29-51.
- Chickering, A., & Braskamp, L. A. (2009). Developing a global perspective for personal and social responsibility. *Peer Review*, 11, 27-30.
- Cummings, W. (2001). *Current challenges of international education* (No. ERIC Document Reproduction Service No. ED464523). Washington, D.C.
- Curran, K. (2003). *Global competencies that facilitate working effectively across cultures*, 2003, from <http://content.monster.com.sg/management/5808/>
- Deardorff, D. (2004). Internationalization: In search of intercultural competence. *International Educator*, 8(2), 13-15.
- Draney, K., & Wilson, M. (2008). A LLTM approach to the examination of teachers' ratings of classroom assessment tasks. *Psychology Science*, 50(3), 417-432.
- Engberg, M. E., & Hurtado, S. (in press). Developing pluralistic skills and dispositions in college: Examining racial/ethnic group differences. *Journal of Higher Education*.
- Fantini, A. (2009). Assessing intercultural competence: Issues and tools. In D. K. Deardorff (Ed.), *The SAGE handbook of intercultural competence*. Thousand Oaks, C.A.: Sage.
- Ferguson, G. A. (1942). Item selection by the constant process. *Psychometrika*, 7(1), 19-29.
- Fischer, G. H., & Parzer, P. (1991). An extension of the rating scale model with an application to the measurement of change. *Psychometrika*, 56(4), 637-651.
- Gliozzo, C. (2002). An agenda for international education. *IIE Networker*, 17.
- Godbey, G. (2002). *Personal communication*. Allentown, PA.
- GPI: <http://www.gpi.hs.iastate.edu/>
- Green, M. (2000). *Toward comprehensive internationalization: An ACE initiative for undergraduate education*. Washington, D.C.: American Council on Education.
- Hammer, M. (2012). The intercultural development inventory: A new frontier in assessment and development of intercultural competence. In M. Vande Berg, R. M. Paige, & K. H. Lou (Eds.), *Student learning abroad* (Ch. 5, pp. 115-136). Sterling, V.A.: Stylus Publishing.
- Hill, D. J. (1991). *Global education and the study abroad program*. Ohio: Renaissance Publications.
- Hunter, W. (2004). *Knowledge, skills, attitudes, and experiences necessary to become globally competent* (pp. 130-131). Washington, D.C.: American Council on Education.
- Hurtado, S. (2003). *Preparing college students for a diverse democracy: Final report to the U.S. Department of Education, OERI, Field Initiated Studies Program*. Ann Arbor, M.I.: Center for the Study of Higher and Postsecondary Education.
- Larry, A. B. (2011). Fostering global perspective taking at American colleges and universities. *Journal of College and Character*, 12(1).
- Lawley, D. N. (1943). The application of the maximum likelihood method to factor analysis. *British Journal of Psychology, General Section*, 33(3), 172-175.
- Lohmann, J. R., Rollins, H. A., & Joseph Hoey, J. (2006). Defining, developing and assessing global competence in engineers. *European Journal of Engineering Education*, 31(01), 119-131.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149-174.
- Masters, G. N. (1988). The analysis of partial credit scoring. *Applied Measurement in Education*, 1(4), 279-297.

- Masters, G. N. (1988). Item discrimination: When more is worse. *Journal of Educational Measurement*, 25(1), 15-29.
- Masters, G. N., & Evans, J. (1986). Banking non-dichotomously scored items. *Applied Psychological Measurement*, 10(4), 355-367.
- Merryfield, M. (1995). *Teacher education in global and international education* (No. ERIC Document Reproduction Service No. ED384601). Washington, D.C.
- Merryfield, M. M. (2001). Moving the center of global education: From imperial world views that divide the world to double consciousness, contrapuntal pedagogy, hybridity, and cross-cultural competence. *Critical issues in social studies research for the 21st century*, 179-208.
- Merryfield, M. M., Lo, J. T. Y., Po, S. C., & Kasai, M. (2008). Worldmindedness: Taking off the blinders. *Journal of Curriculum and Instruction*, 2(1), 6-20.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *ETS Research Report Series*, 1992(1), 1-30.
- Mosier, C. I. (1940). Psychophysics and mental test theory: Fundamental postulates and elementary theorems. *Psychological Review*, 47(4), 355-366.
- Mosier, C. I. (1941). Psychophysics and mental test theory (II. The constant process). *Psychological Review*, 48(3), 235-249.
- Ramsay, J. O. (1991). Kernel smoothing approaches to nonparametric item characteristic curve estimation. *Psychometrika*, 56(4), 611-630.
- Rasch, G. (1960 (1980)). *Probabilistic models for some intelligence and attainment tests* (Reprint, with a Foreword and Afterword by Benjamin D. Wright). Chicago: University of Chicago Press.
- Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In *Proceedings of the fourth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 4, pp. 321-333). Berkeley, C.A.: University of California Press.
- Rasch, G. (1967). An informal report on a theory of objectivity in comparisons. In L. J. Th. van der Kamp, & C. A. J. Vlek (Eds.), *Measurement theory. Proceedings of the NUFFIC International Summer Session in Science in "Het Oude Hof"*, The Hague, July 14-2'8, 1966, University of Leiden.
- Rasch, G. (1968, September). A mathematical theory of objectivity and its consequences for model construction. Paper presented at the *European Meeting on Statistics, Econometrics and Management Science*, Amsterdam, The Netherlands.
- Richardson, M W. (1936). The relation between the difficulty and the differential validity of a test. *Psychometrika*, 1(2), 33-49.
- Rimer, S. (2004, April 27). *Committee urges Harvard to expand the reach of its undergraduate curriculum*. New York Times. Retrieved May 10, 2004, from <http://www.nytimes.com>.
- Reimers, F. M. (2009). Global competency. *Harvard International Review*, 30(4), 24-27.
- Samejima, F. (1968). Estimation of latent ability using a response pattern of graded scores. *ETS Research Bulletin Series*, 1968(1), 1-169.
- Samejima, F. (1972). A general model for free-response data. *Psychometrika*, 37(1), 1-68.
- Samejima, F. (1995). Acceleration model in the heterogeneous case of the general graded response model. *Psychometrika*, 60(4), 549-572.
- Scalise, K., Madhyastha, T., Minstrell, J., & Wilson, M. (2010). Improving assessment evidence in e-learning products: Some solutions for reliability. *International Journal of Learning Technology*, 5(2), 191-208.
- Schwartz, R. A. (2012). The development and psychometric modeling of an embedded assessment for a data modeling and statistical reasoning learning progression (Doctoral dissertation, University of California, Berkeley).
- Sutton, M. (1999). Global education and national interest: The last fifty years. *International Journal of Social Education*, 13, 1-18.
- The American Council on International Intercultural Education Conference (November 15-17, 1996). Educating for the global community, a framework for community colleges. Paper presented at the *Stanley Foundation and The American Council on International Intercultural Education Conference*, Warrenton, V.A.
- The American Council on the Teaching of Foreign Languages (ACTFL). (n.d.). *Test of Oral Proficiency*. <http://www.actfl.org/professional-development/certified-proficiency-testing-program>
- Wilson, M. S., & Dalton, M. A. (1997). *Understanding the demands of leading in a global environment: A first step*. Issues and Observations.
- Wright, B. D., Linacre, J. M., Gustafson, J. E., & Martin-Lof, P. (1994). Reasonable mean-square fit values. *Rasch measurement transactions*, 8(3), 370.
- Wu, M., Adams, R., & Wilson, M. (2007). ACER ConQuest version 2.0: Generalised item response modelling software. Camberwell, Vic: ACER Press.
- Yu, M. N. (2009). Item response theory. Psychological Publishing Co., Ltd., Taipei (in Chinese).