

Reliability and Validity Evidence of Instrument Measuring Competencies for Superior Work Performance

Nornazira Suhairom, Aede Hatib Musta'amal,
Nor Fadila Mohd Amin and Adibah Abdul Latif

Introduction

Economic growth and current development in the field of hospitality and tourism industry have impacted culinary as one of the industry's important niche area. There are increased demands for high-skilled and competent culinary professionals working in the restaurant, catering, and hotels' sector. However, the high turnover rate of culinary professionals in the industry was one of the most confounding employment issues for the culinary profession. Additionally, recent studies have shown that the culinary graduates and young chefs are lack of knowledge, skills and abilities in performing their job. Thus, this study concerns on measuring the mastery level of competencies required for successful career in culinary profession. It is beneficial to have a comprehensive competency measurement instrument developed specifically for the profession can be used to provide data on competencies of current employees that we have in the industry. The establishment of a genuinely valid competency-based assessment approach can yield great benefit, not only to the professions, but to the whole community (Greenstein 2012; Gonczi et al. 1993).

Assessment should be precise, technically sound, producing accurate information for decision making in all circumstances (Dubois and Rothwell 2000; Stetz and Chmielewski 2015). The utilization of survey technique is communal among social science researchers as most of the data collection emphasizes on self-reported data. Nevertheless, the credibility of the instrument as a reliable and valid measurement tool is somehow flouted. It is important to consider the fact that identifying reliability and validity of an instrument is crucial for maintaining the accuracy of the instrument. In order to improve the survey instrumentation, Rasch measurement model was utilized to employ a data-driven model which is designed to measure

N. Suhairom (✉) · A.H. Musta'amal · N.F.M. Amin · A.A. Latif
Faculty of Education, Universiti Teknologi Malaysia, UTM Johor, Malaysia
e-mail: Nornadnazira@gmail.com

culinary professionals' self-assessment on their competencies. The Rasch measurement model was opted for this study because of its sophisticated approach to evaluate patterns of items responses and scale, and item performance (Linacre 2002; Bond and Fox 2007, 2015). Analysis using Rasch measurement model is a more sophisticated approach to evaluate patterns of items responses and scale, and item performance (Chen et al. 2014).

Methodology

The instrument testing was conducted among culinary professionals in 20 commercial kitchens of the 4-star and 5-star hotels in Peninsular Malaysia. The SC-SAT instrument was tested on 111 culinary professionals using a survey which had been conducted for three months.

Data Analysis and Findings

This section describes the data analysis and findings from the instrument testing. The questionnaire was analyzed using Winsteps software, software based on Rasch measurement model for reliability and validity test.

Demographic Profile

There are 111 respondents from 20 hotels in Peninsular Malaysia involved in the study. Table 1 shows the demographic profile of the respondents. About 61 respondents work with the 5-star hotels and 50 respondents work with the 4-star hotels. Majority of the respondents are Malay (83.8 %), followed by Chinese (9 %), Indian (4.5 %), and others (2.7 %).

According to gender classification, there are 74 (66.7 %) male and 37 (33.3 %) female culinary professionals. Among these 111 respondents, 52.3 % of them holds managerial level position (Chef de Partie post and above ranks) while 47.7 % works at nonmanagerial level (Commis and Cook). For each hotel's management, the title of job positions are varies; there are 16 types of job titles among respondents of the study. In terms of educational background, majority of them were Diploma holders (53.2 %), followed by high school graduates (43.2 %). There are three respondents who have Degree and one respondent who has a postgraduate educational accomplishment. For methods of culinary training and education attainment, 66 respondents (59.5 %) reported that they learnt culinary from culinary schools or institutions. Another 40.5 % of the respondents learnt culinary through

Table 1 Respondents' demographic profile

Demographic factors	Factors	f	%
Age (years old)	18–25	31	27.9
	26–35	57	51.4
	36–45	20	18.0
	<46	3	2.7
Job position	Executive Chef	3	2.9
	Executive Sous Chef	1	0.9
	Sous Chef	9	8.1
	Junior sous Chef	5	4.5
	Head Chef	4	3.6
	Pastry Chef	3	2.7
	Banquet Chef	1	0.9
	Demi Chef	10	9.0
	Chef de Partie	13	11.7
	Junior Chef de Partie	4	3.6
	Supervisor	1	0.9
	Chef	2	1.8
	Chef trainer	1	0.9
	Kitchen coordinator	1	0.9
	Cook	6	5.4
Commis	47	42.3	
Culinary experience	Below 5 years	29	26.1
	5–10 years	45	40.5
	11–15 years	19	17.1
	16–20 years	8	7.2
	21 years and above	10	9.0

experiences. About 83.8 % of the respondents have experience working in foreign countries. A large percentage of the respondents (84.7 %) do not have the MOSQ certification.

Person and Item Reliability and Separation Index

In the third instrument testing, the value for person reliability is 0.99 with person separation index of 8.78. Person reliability interpretation is equivalent with Alpha Cronbach (KR-20), which is 0.99. The person separation index value of 8.78 demonstrates that there are 9 levels of person ability that can be categorized in the instrument. With 111 person measuring 159 items in the SC-SAT instrument, Table 2 shows the value of item reliability is 0.94 with separation index of 4.02.

Table 2 Items reliability and separation index for each constructs in SC-SAT

Constructs	Total items	Item reliability	Separation index
Technical	64	0.95	4.15
Nontechnical	52	0.85	2.38
Personal quality	20	0.88	2.72
Physical state	3	0.73	1.66
Self-concept	6	0.70	1.54
Motives	14	0.96	4.80

The finding demonstrates that the probability of the SC-SAT instrument reliability when given to another group of sample with the same characteristics is 0.94. The separation index of 4.02 means that items in the SC-SAT can be categorized into four levels of difficulty. Table 3 shows the value of item reliability and separation obtained for the six constructs in SC-SAT. From the table, it can be seen that most of the constructs showed item reliability value that is greater than 0.70, ranging from 0.70 to 0.96. Physical state and self-concept shows the item separation index below 2 (1.66 and 1.54).

Based on Table 3, all of the constructs are accepted because the item separation indexes are equal to and higher than 2, which is considered as acceptable values except for physical state construct which need to be revised as the value of item separation is 1.72. However, the person reliability for physical state construct is 0.75 indicates a satisfying condition for further analysis.

Item Polarity

Based on Table 4, all of the correlation coefficient is positive for each of the constructs, showing the item ability to measure the competencies is valid (Linacre 2002). There are no items that need to be dropped based on polarity requirement because items are moving in one direction with the constructs.

Table 3 Person reliability and separation index for each constructs in SC-SAT

Constructs	Total items	Person reliability	Separation index
Technical	64	0.97	6.17
Nontechnical	52	0.96	5.14
Personal quality	20	0.89	2.89
Physical state	3	0.75	1.72
Self-concept	6	0.85	2.41
Motives	14	0.88	2.69

Table 4 Polarity of items

Constructs	PTMEA corr				Total items
	Min*	Item	Max*	Item	
Technical	0.47	OPS4	0.87	CRE2	64
Nontechnical	0.53	MGM5	0.74	MGM2	52
Personal quality	0.50	PS13	0.71	PS3	20
Physical state	0.85	PHY3	0.93	PHY2	3
Self-concept	0.79	SC1	0.87	SC3	6
Motives	0.55	MOT14	0.75	MOT9	14

Max* Maximum value; Min* Minimum value

Fit Statistics

Table 5 shows the summary of analysis of Item Fit and Person Fit for the instrument testing. Based on the table, ten respondents were identified to be the misfit person in measuring the six constructs of competency. They are person ID88, ID104, ID112, ID110, ID76, ID41, ID49, ID0, ID64, and ID40.

Accordingly, Table 6 shows the detailed analysis of Person Fit. The analysis shows that these people do not meet the requirement of Rasch model in analyzing the fit characteristics. Thus, suggested these people supposed to be removed from the analysis.

Further, Table 7 shows the item misfit for each of the items in the SC-SAT instrument. Nine items was found to have Infit MNSQ above 1.4 and ZSTD above 2. There is only one item with value of Infit MNSQ below 0.6 and ZSTD value below -2.00; which is SVC5 (I can apply stalls arrangement concept).

Figure 1 depicts the visual presentation of the bubble charts generated by the Rasch Analysis. Item that fits the models' expectations are located in the acceptable values between -2.0 and +2.0. Items which located on the right (>+2.0) are too erratic to be useful whereas items on the left (<-2.0) are too good to be true.

Table 5 Item and person fit for SC-SAT

Constructs	Item				Person			
	Infit MNSQ		Outfit MNSQ		Infit MNSQ		Outfit MNSQ	
	Min	Item	Max	Item	Min	ID	Max	ID
Technical	0.57	SVC5	1.35	OPS2	0.17	ID88	2.55	ID41
Nontechnical	0.64	CAR7	1.90	MGM5	0.06	ID104	3.11	ID49
Personal quality	0.67	PS1	2.02	PS13	0.09	ID104	3.09	ID49
Physical state	0.77	PHY2	0.89	PHY3	0.02	ID112	3.96	ID04
Self-concept	0.74	SC3	1.27	SC6	0.03	ID110	4.65	ID64
Motives	0.76	MOT5	1.48	MOT8	0.20	ID76	5.04	ID40

Table 6 Analysis of person misfit SC-SAT

Measure	Model SE	Infit		Outfit		PTMEA corr	Person
		MNSQ	ZSTD	MNSQ	ZSTD		
0.40	0.12	2.53	9.4	2.54	9.5	0.41	ID40
1.33	0.13	2.33	8.5	2.32	8.5	0.39	ID01
1.43	0.13	2.17	7.8	2.17	7.8	0.39	ID25
0.53	0.12	2.03	6.9	2.06	7.1	0.41	ID49
1.44	0.13	2.04	7.0	2.03	7.0	0.39	ID27
-0.24	0.11	1.90	6.3	1.87	6.1	0.43	ID65
0.50	0.12	1.87	6.0	1.80	5.6	0.41	ID57
2.17	0.13	1.81	6.0	1.80	6.0	0.38	ID41
4.21	0.17	1.80	5.8	1.73	4.8	0.31	ID68
2.03	0.13	1.70	5.3	1.70	5.3	0.38	ID114
0.21	0.12	1.66	4.8	1.61	4.5	0.42	ID38
1.88	0.13	1.62	4.7	1.64	4.8	0.38	ID48
<i>Better fitting omitted</i>							
1.52	0.13	0.39	-7.1	0.38	-7.2	0.39	ID73
2.00	0.13	0.28	-9.3	0.28	-9.3	0.38	ID60
0.00	0.11	0.27	-9.0	0.28	-8.9	0.42	ID30
1.93	0.13	0.24	-9.9	0.24	-9.9	0.38	ID99
1.80	0.13	0.20	-9.9	0.19	-9.9	0.38	ID104
1.91	0.13	0.20	-9.9	0.19	-9.9	0.38	ID75

Table 7 Analysis of item misfit for SC-SAT instrument

Measure	Model SE	Infit		Outfit		PTMEA corr	Item
		MNSQ	ZSTD	MNSQ	ZSTD		
0.34	0.15	1.78	4.6	1.85	4.9	0.60	PS13
0.82	0.15	1.69	4.1	1.78	4.6	0.61	MGM5
0.72	0.15	1.74	4.4	1.73	4.4	0.61	MOT8
-0.56	0.16	1.67	4.2	1.58	3.4	0.56	SC6
-0.64	0.16	1.51	3.3	1.62	3.5	0.56	PS2
0.78	0.15	1.54	3.4	1.59	3.7	0.61	ENT1
0.74	0.15	1.48	3.1	1.45	2.9	0.61	MOT7
-0.01	0.15	1.48	3.1	1.42	2.7	0.58	CAR15
0.29	0.15	1.47	3.0	1.41	2.7	0.59	PHY2
<i>Better fitting omitted</i>							
0.56	0.15	0.56	-3.7	0.57	-3.6	0.60	SVC5

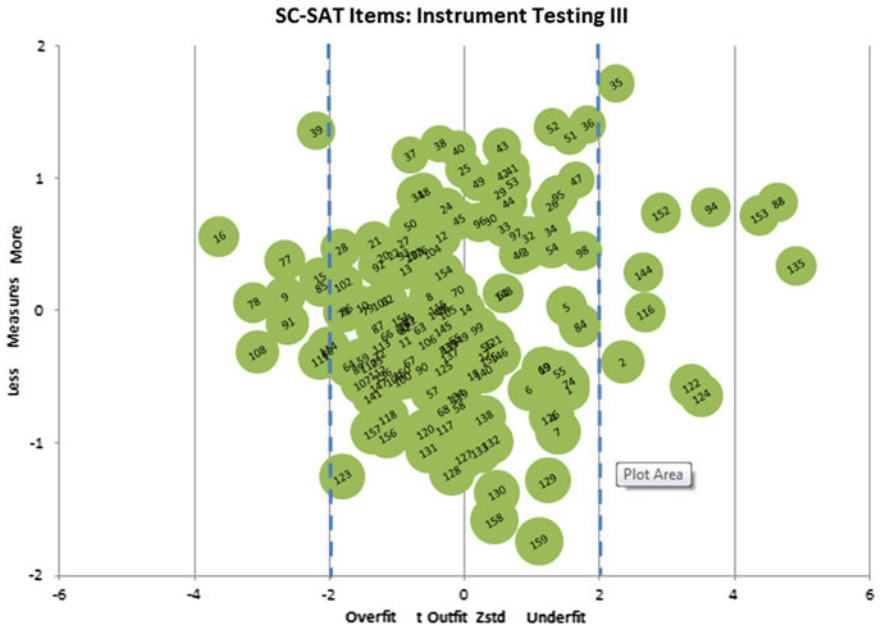


Fig. 1 Visual presentations of fit (quality control) for SC-SAT instrument

Item Dimensionality

Based on Table 8, the raw variance explained by measures is 41.8 %, whereas the unexplained variance in first contrast is 5.5 %.

Table 9 shows the value of raw variance explained by measures and the value of unexplained variance in first contrast for each constructs. The raw variance explained by measures ranging from 42.6 % (nontechnical) to 74.7 % (physical state). It is observed that this value is above the Rasch measurement requirement where the value must exceed 40 %. The range for the unexplained variance in first contrast is 7.1 % (nontechnical) to 13.3 % (physical state) also considered as an acceptable value below 15 %.

Table 8 Standardized residual variance (in eigenvalue)

Condition	Empirical (%)	Modeled (%)
Raw variance explained by measures	41.8	41.7
Unexplained variance in first contrast	5.5	14.3

Table 9 Standardized residual variance (in eigenvalue) for each constructs

Constructs	No. of items	Raw variance explained by measures (%)	Unexplained variance in 1st contrast (%)
Technical	64	47.1	8.3
Nontechnical	52	45.6	7.1
Personal quality	20	42.6	9.3
Physical state	3	74.7	13.3
Self-concept	6	64.0	10.6
Motives	14	47.9	11.7

Standardized Residual Correlation

The largest standardized residual correlation that is used to identify dependent items is displayed in Table 10. There are ten pairs of items that need to be revised because the value is more than 0.70, meaning that these items are highly correlated with each other.

Item Calibration

Analysis on item calibration was done to investigate whether appropriate rating scales are applied. The observed count is the number of times the category was selected across all items and persons. Based on Table 11, the scale used in the questionnaire is 5-point scale which described as 1: Not at All True of Me, 2: Slightly True of Me, 3: Moderately True of Me, 4: Very True of Me, 5: Completely True of Me. The scale number 4 “Very True of Me” is the most selected response from the respondents (48 %). The least response is for scale 1 (Not at All True of Me) with 0 % responses.

The observed average is normal and improved from negative to positive index. The index value starts from -0.38 to $+3.26$ logit. The category probability curve is shown in Fig. 2. Bond and Fox (2007) claimed that each of the rating categories should have a distinct peak in the probability curve graph. However, it can be seen that not the entire peak is clearly seen.

Further analysis on the calculation of Structure Calibration shows the values are not acceptable according to the requirement of $1.4 < SC < 5$ where $[-1.55 - (-2.42) = 0.87]$; thus collapsing is required between scale 1 and scale 2. After collapsing is done, the new Structure Calibration is improved. The observed average is increasing steadily and consistent as shown in Table 12.

The observed average, the average of logit positions modeled in the category is normal and enhanced from negative to positive index. The index value starts from

Table 10 Standardized residual correlations

Corr value	Item	Statement	Item	Statement
0.80	FIN4	I can write menu complete with financial management record	FIN5	I have the ability in implementing labor cost controls for my department
0.77	SCI3	I know the purpose of using food additives in food preparation	SCI4	I know the exact amount food additives to be used in cooking
0.73	OPS4	I know various cutting techniques in preparing foods	OPS6	I know how to apply appropriate cooking methods
0.72	HYG1	I have knowledge on the importance of hygiene for kitchen premises	HYG2	I comply to hygiene rules while in food handling and preparation
0.72	PHY1	I am fit for handling long hours events	PHY2	I am fit for handling heavy equipment in the kitchen
0.71	NUT2	I am skilled at developing menu based on special dietary needs	NUT3	I know the emergence of health products such as organic foods and genetically modified foods
0.70	OPS9	I have knowledge of producing products with original flavors	OPS 10	I am skilled at deploying size and portion of the food products
0.70	CAR13	I engage in activities that are directly linked to my performance	CAR14	I keep informed on affairs, structures, and processes in my profession
0.69	SVC3	I can identify type of buffet display and setting for catering services	SVC4	I can apply method of buffet display and setting dishes arrangement
0.68	INO3	I incorporate new ingredients to the recipe	INO4	I incorporate change in traditional cooking method

Table 11 Observed average at 5-point scale (12345)

SUMMARY OF CATEGORY STRUCTURE. Model="R"										
CATEGORY LABEL	OBSERVED SCORE	OBSVD COUNT	%	AVRGE	SAMPLE EXPECT	INFIT MNSQ	OUTFIT MNSQ	STRUCTURE CALIBRATN	CATEGORY MEASURE	
1	1	74	0	-.38	-.57	1.11	1.17	NONE	(-3.75)	1
2	2	638	4	.09	.05	1.03	1.03	-2.42	-2.05	2
3	3	4587	26	.77	.83	.94	.94	-1.55	-.34	3
4	4	8476	48	1.87	1.82	.94	.94	.69	2.01	4
5	5	3873	22	3.26	3.30	1.06	1.05	3.28	(4.43)	5

OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.

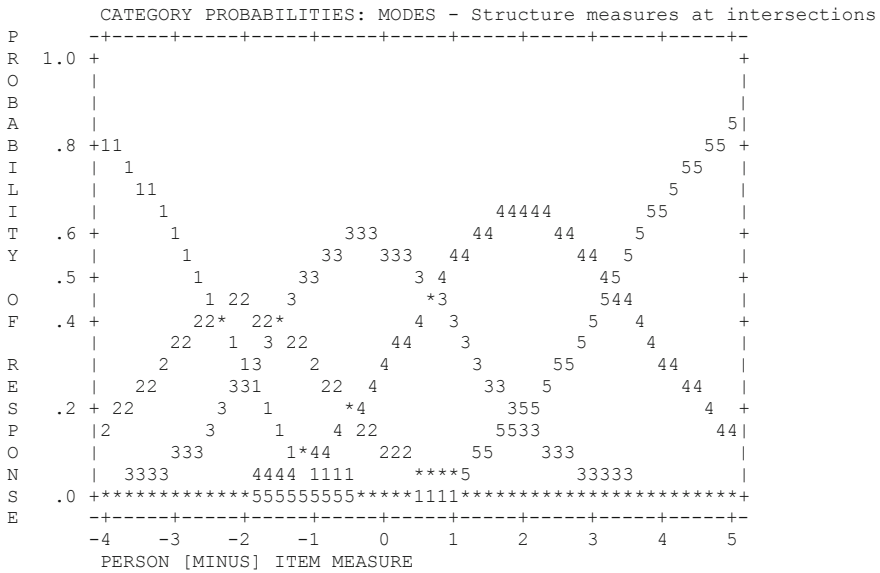


Fig. 2 Category probability curve at 5-point scale (12345)

Table 12 Observed average at 4-point scale!MediaObject ID="MO19">

SUMMARY OF CATEGORY STRUCTURE. Model="R"									
CATEGORY LABEL	OBSERVED SCORE	OBSVD COUNT	SAMPLE %	INFINIT	OUTFIT	INFINIT MNSQ	OUTFIT MNSQ	STRUCTURE CALIBRATN	CATEGORY MEASURE
1	1	712	4	-0.80	-0.86	1.04	1.04	NONE	(-3.49)
2	2	4587	26	-0.08	-0.02	.94	.94	-2.31	-1.25
3	3	8476	48	1.04	1.00	.94	.94	-0.15	1.19
4	4	3873	22	2.45	2.48	1.06	1.05	2.46	(3.61)

OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.

-0.80 to +2.45 logit, demonstrating that it is increased by category value. The new category probability curve is shown in Fig. 3.

It can be observed from the figure that the entire peak can be seen distinctively. Further analysis on the calculation of Structure Calibration shows the values are well accepted according to the requirement of $1.4 < SC < 5$ where $[-0.15 - (-2.31) = 2.16]$. In an attempt to revise the categorization, the item and person reliability and separation index is reanalyzed for the calibrated scale of 11234. Table 13 shows the comparison of separation index value before and after scale calibration. Result shows that after category collapsing was done, the value of item separation index is increased from 4.02 to 4.03. On the other hand, the value of person separation index is maintained at 8.78. The value of mean person decreases from 1.81 to 0.98 (standard deviation 1.30).

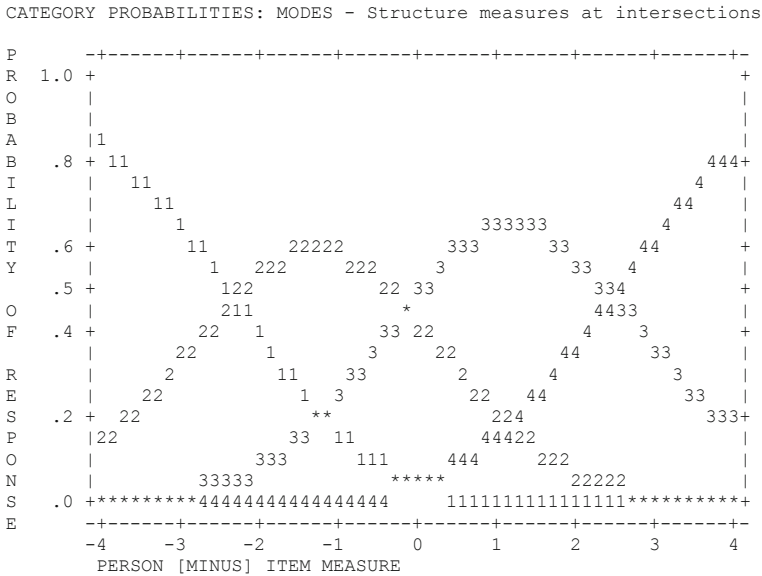


Fig. 3 Category probability curve at 4-point scale (1234)

Table 13 Comparison before and after scale calibration

Condition	Item		Person		Person mean
	Reliability	Separation	Reliability	Separation	
Before scale calibration	0.94	4.02	0.99	8.78	1.81
After scale calibration	0.94	4.03	0.99	8.78	0.98

The utilization of 5 likert-type rating scale is suggested for the next stage of instrument testing after taking into account that the changes in person and item separation appear to be meaninglessly small.

Differential Item Functioning (DIF)

It is crucial that the items in the designated SC-SAT instrument should not advantage (or disadvantage) culinary professionals from different groups. The differential item functioning (DIF) analysis is conducted to strengthen the psychometric evaluation of the instrument. The major purpose of DIF analysis is to identify whether there are biases exists among items in the SC-SAT instrument from the aspects of gender. To analyze DIF, Winstep perform the two-tailed t-test

Table 14 DIF analysis of SC-SAT items

Person class	DIF measure	Person class	DIF measure	DIF contrast	<i>t</i>	Item
1	-0.63	2	0.04	-0.67	-2.05	OPS2
1	-0.24	2	0.43	-0.67	-2.10	SVC3
1	1.95	2	1.27	0.68	2.34	SCI1
1	1.75	2	0.75	1.01	3.37	SCI2
1	1.40	2	0.75	0.65	2.17	SCI3
1	1.09	2	1.71	-0.62	-2.13	FIN4
1	0.70	2	0.04	0.66	2.07	QUA2
1	-0.27	2	0.87	-1.14	-3.63	EMO2
1	-0.63	2	0.04	-0.67	-2.05	EMO5
1	-0.84	2	-0.03	-0.82	-2.45	EMO6
1	0.42	2	1.10	-0.68	-2.23	ENT3
1	-0.31	2	-1.03	0.72	2.12	SC6
1	-0.31	2	-1.27	0.95	2.78	PS12
1	0.93	2	-0.88	1.81	5.52	PS13
1	-0.20	2	-1.03	0.84	2.47	PS14

1 Male, 2 Female

analysis to test the significant difference between the difficulty indexes. In DIF analysis, the cut off point is the critical *t* value within range of $+2.0 \geq t \geq -2.0$ and $+0.5 \geq \text{DIF contrast} \geq -0.5$ at 95 % confidence level.

Items which have DIF contrast value outside the range $\geq +0.5$ or ≤ -0.5 need to be revised after considering the *t* value. Results for the DIF Analysis of SC-SAT items based on gender are presented in Table 14. There are fifteen items which is detected as items that have bias between the two groups of male and female culinary professionals in the SC-SAT instrument. The analysis shows that most of the DIF measure for Person Class 1 (male) is smaller than DIF Measure for Person Class 2 (female), indicating that male culinary professionals more easy to endorse their self-reflections towards the competency items.

Item Targeting

The data were delved further to determine the Malaysian Culinary Professional's Competency Profile based on the SC-SAT instrument. Additional analysis were conducted to demonstrate the ability of the Rasch Analysis diagnoses in constructing the competency profiling based on the item difficulty and person ability. The heart of the Rasch Analysis is presented in Fig. 4, where the map of the person and items was displayed in tandem. The mean for all items are indicated as "M" (Item Mean) starts at 0.00 logit while the Person Mean (also marked as "M") is

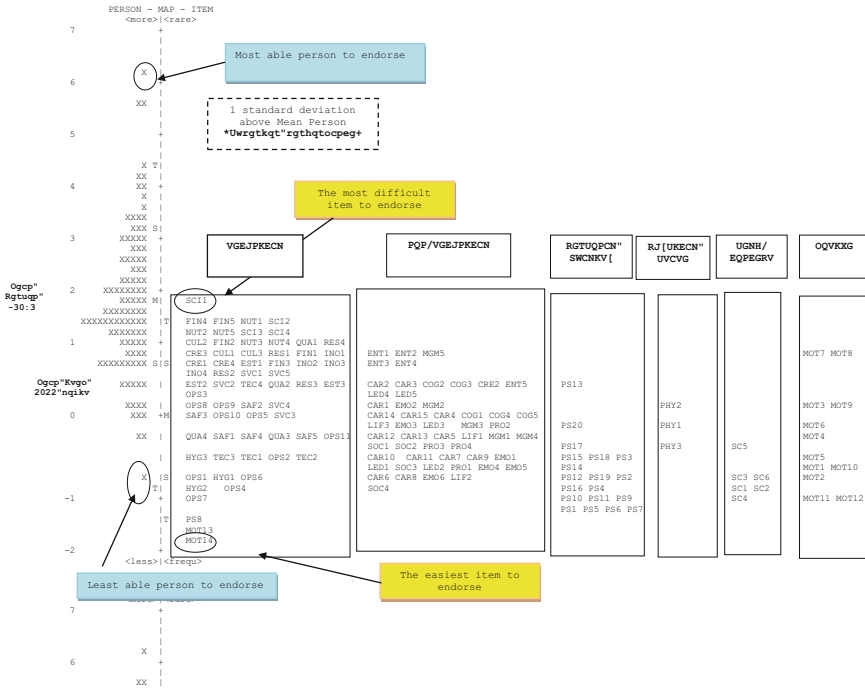


Fig. 4 Item-person wright map based on the SC-SAT

observed at +1.81. “S” is one standard deviation away from the mean, whereas “T” is two standard deviations away from the mean. From the Item-Person map, it shows that the Person Mean is above the Item Mean. Respondents’ ability was arranged according to ascending order from the lowest to the highest ability in performing the items.

As shown in Table 15, the most difficult item is located at +1.72 logit and the easiest item is located at -1.74 logit with the standard deviation of 1.29, inferring the small spread within the data. Though the items still targeting at groups of person with moderate ability and below, there is an even distribution of persons according to their abilities along the logit scale. This shows that there is a slight improvement on item targeting. The most difficult item that respondents gave endorsement is item SC11 from constructs technical competency “knowledge of cooking chemistry”.

Table 15 Item difficulty level and person response level

	Item difficulty level		Person response level	
	Maximum	Minimum	Maximum	Minimum
Mean	0.00		+1.81	
Entry number	SC11	MOT14	ID83	ID46
Logit value	+1.72	-1.74	+6.18	-0.62

Item M14 is the easiest items from motive construct “career as Chef brings the utmost satisfaction”. There are 21 off target items with no respondents, which mean that these items are too easy. The maximum logit for person is +6.18 logit which is represented by person ID83, followed by two people at +5.67 logit (Person ID10 and ID82). The minimum logit for person is -0.62 logit (Person ID46).

From Fig. 4, there are two categories of items spanned along the positive and negative logit scale. There are 72 items above the mean (45 %) and another 87 items below the mean (55 %). This shows only 45 % of positive person response level, showing that these percentages of the respondents have perceptions to agree with the items. Accordingly, the findings demonstrate that they are capable in carrying out the competencies. Person ID83 who demonstrate the highest logit value (+6.18 logit) is a male, Chinese person who holds job position as a Sous Chef in a 5-Star hotel. He has been holding the position around 6–10 years. He is aged between 26 and 35 years old with more than 11–15 years of experience in the culinary industry. The person also has experience working in foreign country. Nevertheless, the person does not have MOSQ certification.

Discussion

As discussed earlier, Rasch measurement analysis was initially, conducted with 203 items, and conceptually ordered from low to high level of difficulty. It was concluded that a reliable linear, unidimensional scale of competencies for superior work performance was created using culinary professional views. With such detailed precision, these results mean that valid inferences could be made from the SC-SAT instrument. Since the scale data were shown to be reliable, valid inferences were drawn from the scale. Findings from the study have shed lights on the construct validity of the scales constructed. The study emphasized on six aspects of Rasch Analysis diagnoses which are (i) item and person reliability and separation index, (ii) item fit, (iii) item polarity, (iv) item dimensionality, (v) item calibration, and (vi) differential item functioning. The aspect of item targeting and consequently, competency profiling based on the SC-SAT instrument were discussed further. A closer look at the responses given by the culinary professionals in answering the SC-SAT may indicate which aspects specifically are sound and which may need attention to further developed their competence at work. The development of SC-SAT has put forward a better technique of competency measurement that is purposely developed for employees’ professional development.

Assessment should include a spectrum of strategies where the process and products are emphasized. Assessment should be communicated, integrated in a day to day basis, stimulate thinking, build prior knowledge and construct meaning. Assessment results should be routinely revised and provide a proper database (Dubois and Rothwell 2000; Stetz and Chmielewski 2015). Formative assessment should be responsive in a way that it provides opportunities for self-reflections and revision. Hence, the betterment for model of assessment for professional

development is to provide feedback. Workers should receive feedback routinely, recognizing achievement beyond the scores.

Other than culinary profession, studies focusing on aspects of assessment of professional employees' competence and performance, addressing the question of self-assessment, and the means to assure more objective measurements of competence and performance were conducted widely among vocational profession (Winther and Klotz 2013), health professionals (Bashook 2005; Nicholson et al. 2012) information technology professionals (Azrilah et al. 2008), sales professionals (Lambert et al. 2014) and management professionals (Sisson and Adams 2013). The studies also attempts to develop applications of findings in identifying performance at workplace using a bevy of assessment methods.

Conclusion

Analyses of SC-SAT instrument items fully support its function as a useful measure of competencies. All items were analyzed and a minor modification has been made in order to achieve an adequate model fit. Data from the instrument testing provide evidence that the psychometric evaluations of the instrument are improved from one stage to another. The newly developed SC-SAT provides opportunity for culinary professionals to identify and measure their own competencies where the result can be used to identify how well they are doing. SC-SAT is considered as a norm-referenced measurement tool that is expected to possess a high degree of accuracy, discriminating those who perform excellently with those who are low performers, and functioning. Rasch Analysis has assisted the researcher in improving the quality of SC-SAT instrument, providing evidence to support the validity of the SC-SAT instrument. This study will be of better quality by implementing a number of improvements in a certain area such as in improving the item targeting.

References

- Azrilah, A. A., Azlinah, M., Noor Habibah, A., Sohaimi, Z., Hamza, A. G., & Mohd Saidfudin, M. (2008). Development of Rasch-based descriptive scale in profiling information professionals' competency. *ISIT, 2008*, 1–8.
- Bashook, P. G. (2005). Best practices for assessing competence and performance of the behavioral Health workforce. *Administration and Policy in Mental Health and Mental Health Services Research, 32*(5–6), 563–592. doi:10.1007/s10488-005-3265-z.
- Bond, T., & Fox, C. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: LEA.
- Bond, T., & Fox, C. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Mahwah, NJ: LEA.
- Chen, W.-H., Lenderking, W., Jin, Y., Wyrwich, K. W., Gelhorn, H., & Revicki, D. A. (2014). Is Rasch model analysis applicable in small sample size pilot studies for assessing item

- characteristics? An example using PROMIS pain behavior item bank data. *Quality of Life Research* 1–9.
- Dubois, D. D., & Rothwell, W. J. (2000). *The competency toolkit* (Part II). HRD Press.
- Gonczi, A., Hager, P., & Athanasou, J. (1993). The development of competency-based assessment strategies for the professions. *National Office of Overseas Skills Recognition, Australia, Research P.*
- Greenstein, A. (2012). *Assessing 21st century skills*. United States of America: Corwin.
- Lambert, B., Plank, R. E., Reid, D. A., & Fleming, D. (2014). A competency model for entry level business-to-business services salespeople. *Services Marketing Quarterly*, 35, 84–103.
- Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions*, 16(2), 878.
- Nicholson, P., Griffin, P., Gillis, S., Wu, M., & Dunning, T. (2012). Measuring nursing competencies in the operating theatre: Instrument development and psychometric analysis using item response theory. *Nurse Education Today*, 6–11. doi:[10.1016/j.nedt.2012.04.008](https://doi.org/10.1016/j.nedt.2012.04.008)
- Sisson, L. G., & Adams, A. R. (2013). Essential hospitality management competencies: The importance of soft skills. *Journal of Hospitality & Tourism Education*, 25(3), 131–145. doi:[10.1080/10963758.2013.826975](https://doi.org/10.1080/10963758.2013.826975).
- Stetz, T. A., & Chmielewski, T. L. (2015). *Competency modeling documentation* (pp. 1–22).
- Winther, E., & Klotz, V. K. (2013). Measurement of vocational competences: An analysis of the structure and reliability of current assessment practices in economic domains. *Empirical Research in Vocational Education and Training*, 5(1), 2. doi:[10.1186/1877-6345-5-2](https://doi.org/10.1186/1877-6345-5-2).