

Item response theory-based validation of Taiwanese patient safety culture measurement instrument

Abstract

The Chinese version of the Safety Attitudes Questionnaire (SAQ-C) was developed and tested in 2007. It consists of 5 domains: teamwork climate (TC, 5 items), safety climate (SC, 6), job satisfaction (JS, 5), perception of management (PM, 10), and working condition (WC, 4). The issue arose that PM had too many items because the same 5 items were asked twice: one for unit management and the other for hospital management. Unfortunately, in many Asian countries, such a management level classification does not provide useful information. Thus, each pair was collapsed into one as “management in this work setting”. In addition, 2 overly general items were dropped. Thus, PM ultimately had 4 items. The new version of SAQ-C became a very compact 24-item instrument named the Taiwanese Patient Safety Culture Survey (TPSC).

We then validated TPSC, but we took a different road. Thus far, almost all survey instrument validations in healthcare have been done using the linear regression-based confirmatory factor analysis (CFA). However, survey responses more often than not use a Likert scale, which is ordinal not interval (continuous). Therefore, the habitually used CFA has not been correctly used, either because of negligence or ignorance. For a scientifically sound analysis, we used the multidimensional item response theory (MIRT) that takes care of this issue. To check the model fit, we used limited-information goodness-of-fit tests with M_2^* statistics considering the sparse contingency table.

In the first quarter of 2009, invitation letters to encourage participation in TPSC were circulated among hospitals in Taiwan. From April 1 to December 31, TPSC was administered to healthcare professionals voluntarily. Because the process was paper-based, the returned questionnaires were entered into a computer database manually. A total of 23,999 questionnaires were returned, of which 4,596 were missing the hospital level variable and were dropped because we intended to build a multigroup model by hospital level.

We examined each item and domain and their parameters, such as factor loadings and variance/covariance; all were satisfactory. The overall model, GOF, and M_2^* statistics were investigated. The root means square error of approximation (RMSEA) was 0.03 (cut-off < 0.06), and the non-normed fit index (NNFI, or the Tucker-Lewis index) was 0.98 (cut-off > 0.95), suggesting TPSC is well validated in the MIRT (IRT) framework. In addition, we conducted the albeit flawed CFA, and the model fit was satisfactory too.

Ultimately, we can safely say that TPSC, the downsized version of SAQ-C, is well validated using either the classic linear-regression-based CFA or the newly developed IRT-based approach. This validation study will help researchers perform their own studies with TPSC.

Volume 7 Issue 4 - 2018

Heon-Jae Jeong,¹ Wui-Chiang Lee,² Hsun-Hsiang Liao,³ Cheng-Fan Wen,⁴ Pa-Chun Wang⁵

¹Joint Commission of Taiwan, New Taipei City, Taiwan

²Department of Medical Affairs and Planning, Taipei Veterans General Hospital & National Yang-Ming University School of Medicine, Taipei, Taiwan

³Joint Commission of Taiwan, 5F, No.31, Sec. 2, Sanmin Rd., Banqiao District., New Taipei City, Taiwan

⁴Division of Quality Improvement, Joint Commission of Taiwan, New Taipei City, Taiwan

⁵Joint Commission of Taiwan, New Taipei City, Taiwan, Quality Management Centre, Cathay General Hospital, Taipei, Taiwan

Correspondence: Wui-Chiang Lee, Department of Medical Affairs and Planning, Taipei Veterans General Hospital & National Yang-Ming University School of Medicine, Taipei, Taiwan, Tel: +886-2-28757120, Fax +886-2-28757200, Email leewuichiang@gmail.com

Co-correspondence: Hsun-Hsiang Liao, Joint Commission of Taiwan, 5F, No.31, Sec. 2, Sanmin Rd., Banqiao District., New Taipei City, Taiwan, Tel: 886-2-89643902, Email shing@jct.org.tw

Received: June 22, 2018 | **Published:** July 17, 2018

Introduction

“Everything boils down to culture.” For anyone who works in the field of patient safety, it does not take long to recognize the truth of this statement. Yet this consensus did not emerge at the very beginning of the patient safety era. Rather, we thought that just changing care processes and adding automated systems would suffice in dropping the adverse event rate—probably to zero. However, it did not take much time till such naïve optimism was completely nullified. Many researchers have focused on why such efforts, represented by physical improvement, could not solve problems perfectly. As evidence has grown, healthcare professionals reached an agreement: safety culture is a must-have ingredient to true improvement in patient safety.¹⁻⁴

The evidence was clear: history depicts that even if the very same safety improvement program were implemented, the effectiveness of the program—even to the success or failure level—would largely vary according to the cultural background of the target area, such as the hospital or even country. This suggests that we have to either adjust

the program most appropriately to the cultural background of the place where it would be transplanted or influence the culture so that it accepts the program without resistance, like terraforming.^{5,6}

Figuratively, such a recommendation leads us to the premise that we know and understand the topography of safety culture of the place in which we are interested. Were it not for this information, we could never plan, execute, or evaluate a program precisely. Peter Drucker once said “*If you can't measure it, you can't improve it*”⁷; this definitely applies to cultural issues in patient safety. Once such a need for measurement was recognized, safety enthusiasts developed an instrument without a single “nay”. Among the several safety culture measurement instruments,⁸ one of the most popular in the world is the Safety Attitudes Questionnaire (SAQ). Developed by Bryan Sexton, the original SAQ consists of six domains: teamwork climate (TC), safety climate (SC), job satisfaction (JS), perception of management (PM), stress recognition (SR), and working condition (WC).⁹ It provides views from different angles of safety culture.

Taiwan has been a trailblazer in adopting SAQ. With permission from the original developer Bryan Sexton, Taiwanese researchers translated SAQ into Chinese and developed the SAQ-Chinese version (SAQ-C). While translating, they found two items—one from TC (“In this work setting it is difficult to speak up if I perceive a problem with patient care”) and the other from SC (“In this clinical area, it is difficult to discuss errors”)—that did not work well and were dropped. Their factor loadings were around 0.30, which was probably due to negative questions not performing as intended in Chinese. In addition, the SR domain itself was a completely different animal among the original 6 SAQ domains—not only in the Chinese version, but also in several countries.^{10–12} As Jeong et al. proved using the bifactor model, SR cannot be in the SAQ.¹² Thus, the whole SR domain (4 items) was

discarded. The pilot version of SAQ-C was administered to volunteers in Taiwan, and 45,252 questionnaires were returned. With a very positive results, the instrument officially obtained the name “SAQ-C” in 2008.¹³

Despite its successful validation, SAQ-C was criticized in two ways. First, although SAQ-C utilized a 5-point Likert scale that is an ordinal scale, the scoring formula from the developer treats it as an interval scale (please think of it as continuous scale). This issue can nullify any results from SAQ-C. Second, compared to the other domains, PM contains 10 items, which—practically speaking—is too many and suspected to be less efficient, leading respondents to leave several items missing until eventually the whole questionnaire is scrapped. These items are listed in Table 1.

The goal of this study is clear. We developed a method to treat a Likert scale as an ordinal scale, as we should. We then re-validated the instrument with the new method. In this process, we shrank the PM domain into 4 items for efficiency. As Table 1 denotes, the 4 items were the most representative items among the small tweaked versions of the original 10 items.

To materialize these goals, we utilized item response theory (IRT) graded response model (GRM), which can handle ordinal data and process them, yielding interval scale results. Specifically, multidimensional-IRT (MIRT) was used, because we have 5 domains.^{14–16} In order to distinguish the new instrument from older SAQ-C, we call the new one Taiwanese Patient Safety Culture (TPSC) measurement instrument. Indeed, this is the official name that the Joint Commission of Taiwan uses for this instrument; the contents are found in Table 3.

Methods

Developing TPSC

Through several discussion sessions, an expert group decided

to remove 6 items from the PM domain in addition to the already removed SR domain (4 items), one item from TC and the other from SC domain; thus, TPSC has 12 items fewer than the original SAQ. Table 1 summarizes the modifications in the PM domain. Many Asian countries, including Taiwan, do not distinguish the unit manager and hospital management. Specifically, healthcare professionals seldom have a chance to see the C-suite people. Thus, with the SAQ-C, people had a hard time answering such items. Eventually, we combined these two manager levels into one (see Table 1). In addition, PM5 and 6 in SAQ-C did not give meaningful information; they contained overly general ideas and, thus, only increased the burden for respondents. In sum, TPSC had 24 items across 5 domains (detailed items and the domain list are depicted in Table 3).

Table 1 Item changes in PM domain from SAQ-C to TPSC

SAQ-C (PM)		TPSC (PM)	
ID	Items	ID	Items
PM1	Unit management supports my daily efforts	PM1	Management in this working setting fully supports my daily efforts
PM2	Hospital management supports my daily efforts		
PM3	Unit magagement doesn't knowingly compromise patient safety	PM2	Management in this working setting doesn't knowingly compromise patient safety.
PM4	Hospital magagement doesn't knowingly compromise patient safety		
PM5	Unit management is doing a good job	PM4	The management unit in this work setting deals with problem personnel constructively.
PM6	Hospital management is doing a good job		
PM7	Problem personnel are dealt with constructively by our unit management	PM3	I get adequate, timely information about events in this working setting that might affect my work.
PM8	Problem personnel are dealt with constructively by our hospital management		
PM9	I get adequate, timely info about events that might affect my work from unit management		
PM10	I get adequate, timely info about events that might affect my work from hospital management		

Data collection

In the first quarter of 2009, invitation letters encouraging participation in TPSC were sent out to hospitals in Taiwan. From

April 1 to December 31, TPSC was administered to volunteers. The processes relied on paper-based versions of the questionnaire, so returned questionnaires were entered into a computer database manually.

Model development to validate TPSC

We built a MIRT model following the correlated factor structure, allowing each domain (latent trait) to be correlated. In addition, considering the consensus in Taiwan's medical society—the hospital level is closely related to quality and safety—we added a multigroup structure with the hospital level variable. Readers might think of it as a kind of controlling method, much like a categorical covariate, although technically it is not. This structure is a clear depiction that equality constraints would equivalently spread across all levels of hospitals with participants.¹⁷

Checking the model fit

Generally, the classic Bock-Aitkin's Expectation-Maximization (BAEM) provides various goodness-of-fit (GOF) indices. However, ordinal data structures, like a 5-point Likert scale, do not allow us to harness such popular indices' full-information GOF tests; for

example, χ^2 or G^2 did not work because the contingency table was too sparse.^{18,19} Therefore, we had to use a limited-information GOF test based on M_2^* statistics, although the number of available GOF indices dropped drastically.^{18,20}

All analyses were performed using software packages for item response theory, flexMIRT 3.51 (Vector Psychometric Group, LLC, Chapel Hill, North Carolina).¹⁷

Results

Characteristics of respondents

A total of 23,999 questionnaires were returned, among which 4,596 questionnaires were missing the hospital-level variable. As we were seeking to build a multigroup model, we dropped those questionnaires. Table 2 summarizes the characteristics of the remaining 19,403 questionnaires used in the MIRT analysis.

Table 2 Characteristics of Respondents

Characteristics	N	%
Gender		
Male	3,303	17.0
Female	16,100	83.0
Age group (years old)		
≤ 20	45	0.2
21-30	8,887	45.8
31-40	7,031	36.2
41-50	2,734	14.1
51-60	649	3.3
> 60	57	0.3
Job Type		
Physicians	2,381	12.3
Nurses	12,204	62.9
Technicians	2,179	11.2
Pharmacists	857	4.4
Administrative staff	846	4.4
Others	936	4.8
Hospital level (N of hospitals)		
Medical centers (13)	11,417	58.8
Regional hospitals (21)	6,619	34.1
District hospitals (10)	1,259	6.5
Psychiatric hospitals (1)	108	0.6
Total	19,403	100.0

Like many other studies administering questionnaire to healthcare professionals, most questionnaires were returned from females, which is not surprising as most nurses are female.^{4,16,21,22} In addition, for each characteristic, the number of respondents in a certain category varies a lot. Female nurses, age 20–40, were the majority in each characteristic; regarding hospital level, 58.8% questionnaires were returned from medical centres, which employ many employees. Some might suggest checking the representativeness of the samples; however, one of the strengths of the IRT framework is that representativeness is not

significantly influenced by respondents' mix.²³

Running the MIRT model and its results

Table 3 describes all item-level parameters. First, we checked factor loadings; each was equal to or higher than 0.67 (TC1), satisfying the generally accepted threshold of 0.5.²⁵ The loadings ranged up to 0.94 (JC3 and JC4). Thus, loadings varied a lot, meaning that a simple mean of domain scores cannot be justified. Therefore, scores should be obtained in the form of factor scores.

Table 3 Item-level parameters from MIRT

ID	Domain: Definition (N. of items)	Factor Loading	Item Parameters				
			a	c1	c2	c3	c4
Teamwork Climate (TC): Perceived quality of collaboration between personnel (5)							
TC1	Staff input is well received in this clinical area.	0.67	1.52	5.04	2.59	0.80	-1.35
TC2	Disagreements in this clinical area are resolved appropriately (i.e., not who is right, but what is best for the patient).	0.74	1.85	5.87	3.56	1.55	-0.56
TC3	I have the support I need from other personnel to care for patients.	0.80	2.24	7.33	4.36	2.32	-0.56
TC4	It is easy for personnel here to ask questions when there is something that they do not understand.	0.80	2.29	7.92	5.28	3.18	0.47
TC5	The physicians and nurses here work together as a well-coordinated team	0.76	1.98	5.92	3.77	1.81	-0.43
Safety Climate (SC): Perception of a strong and proactive organizational commitment to safety							
SC1	I would feel safe being treated here as a patient.	0.81	2.33	7.03	4.53	2.06	-0.62
SC2	Medical errors are handled appropriately in this clinical area	0.81	2.33	7.61	5.19	2.63	-0.14
SC3	I know the proper channels to direct questions regarding patient safety in this clinical area.	0.79	2.21	7.57	5.38	2.76	0.00
SC4	I receive appropriate feedback about my performance.	0.77	2.07	5.58	3.51	1.26	-1.63
SC5	I am encouraged by my colleagues to report any patient safety concerns I may have.	0.70	1.68	6.34	4.09	1.31	-1.16
SC6	The culture in this clinical area makes it easy to learn from the errors of others.	0.67	1.53	5.64	3.77	1.52	-0.98
Job Satisfaction (JS): Positivity about the work experience (5)							
JS1	I like my job.	0.80	2.26	6.85	4.00	1.91	-0.68
JS2	Working here is like being part of a family.	0.89	3.41	9.51	6.77	3.39	-0.19
JS3	This is a good place to work.	0.94	3.39	11.84	8.16	4.00	-0.80
JS4	I am proud to work in this clinical area.	0.94	4.71	12.92	9.34	4.09	-0.39
JS5	Morale in this clinical area is high.	0.88	3.22	8.32	5.47	2.03	-1.59
Perception of Management (PM): Approval of managerial action (4)							
PM1	Management in this working setting fully supports my daily efforts.	0.80	1.64	6.32	3.24	1.40	-1.65
PM2	Management in this working setting doesn't knowingly compromise patient safety.	0.75	1.95	6.96	5.31	2.76	0.48
PM3	I get adequate, timely information about events in this working setting that might affect my work.	0.85	2.75	8.35	5.64	2.46	-1.18
PM4	The management unit in this work setting deals with problem personnel constructively.	0.84	2.66	7.04	4.79	1.69	-1.74
Working Conditions (WC): Perceived quality of the work environment and logistical support (4)							
WC1	The levels of staffing in this clinical area are sufficient to handle the number of patients.	0.64	1.40	3.46	1.87	0.29	-1.81
WC2	This hospital does a good job of training new personnel.	0.79	2.20	6.68	4.56	2.12	-0.62
WC3	All the necessary information for diagnostic and therapeutic decisions is routinely available to me.	0.85	2.73	9.01	4.51	2.46	-1.44
WC4	Trainees in my discipline are adequately supervised.	0.87	2.99	9.55	6.68	3.09	-1.00

Note. (a) discrimination; (c1-c4): intercepts.

As Table 4 indicates, all correlations between the 5 domains are high enough (the lowest, between JS and WC, was 0.77). Although the topic of this article is not calculating individual participant's domain scores, getting them requires considering this variance/covariance matrix; otherwise, we end up ignoring the blueprint of the complex

structure among domains and can only achieve simple mean scores at best.¹²

We examined GOF with M_2^* statistics-based fit indices.^{18,20} The root mean square error of approximation (RMSEA) was 0.03 (cut-

off < 0.06), and the non-normed fit index (NNFI, or the Tucker-Lewis index) was 0.98 (cut-off > 0.95), suggesting TPSC is well validated in the MIRT framework.

Table 4 Variance/covariance matrix (Lower triangle)

	TC	SC	JS	PM	WC
TC	1.00				
SC	0.91	1.00			
JS	0.80	0.81	1.00		
PM	0.84	0.88	0.82	1.00	
WC	0.80	0.85	0.77	0.89	1.00

Discussion

Traditional CFA and MIRT

In addition to the MIRT-based validation, we also conducted a typical linear regression-based CFA without the IRT component and found that most GOF indices were satisfactory, although not described here. **Thus, this article guarantees that TPSC is a validated instrument in both the traditional linear CFA and IRT frameworks.** However, regardless of whether the traditional way was validated, we still suggest utilizing MIRT. The linear-regression based-CFA theoretically cannot handle an ordinal scale like a 5-point Likert scale that TPSC uses, although in the field all too often CFA is used where it should not be. Again, treating a Likert scale as a linear continuous scale is just like using a simple regression for dichotomized data instead of logistic regression. Another reason we favour MIRT is that it produces the finest possible granulation of the results. MIRT provides safety managers with surgical level precision for each respondent's information.

When to revalidate

At this point, readers might question why we conducted a validation study on an already validated instrument. We propose two scenarios. The first is what we described in this article. In order to enhance understanding, we use the following analogy: TPSC (instrument) is fish. Thus far, the fish has lived in seawater (the world of the classic CFA). Now we want to move the fish to freshwater (the IRT paradigm). Before we execute this migration, we must test whether the fish can survive well in the radically different types of water. This is the instrument validation that we showed with MIRT in earlier sections.

Second, imagine a different situation. The fish continued growing until it needed a new fishbowl. We know the relationship between the fish and the fishbowl has changed. There are a number of examples, such as language change, item number change, and contents modification. Technically any change in the instrument is a signal for validation. However, is it possible or even practical to conduct a validation study so often? How big of a change warrants a non-ignorable sign of validation? As always, it ought to be left to the researcher's discretion. The change from SAQ-C to TPSC was definitely in need of validation. However, the combination between what follows and IRT may relax the stringent need for validation.

Classic test theory and IRT

A decade ago, we were sitting in classrooms and taking exams. Many countries still use this method for college entrance exams. If we were in the universe of classic test theory (CTT),²⁵ literally any change would trigger the validation process. In the CTT realm, just a change of sequence between a couple of items leads to revalidation. CTT regards the whole test instrument as a complete kit, so a small change means we develop a new version of the survey. In addition, environmental factors when administering the test, such as noise and modality, must be controlled if this approach is used for an entrance exam. Furthermore, the items cannot be reused, so building an item bank and applying computer-adaptive test are nonsense.^{26,27} In sum, in the world of CTT, no change can slide; pure CTT enthusiasts may claim that even the fonts should be the same.

IRT does not work that way. Of course, instruments should be validated in this realm as well, but the biggest difference between IRT and CTT is that IRT is item oriented, not test oriented. A practical scenario is that item 1 (Table 3) used "nurse" in the original SAQ instead of "staff". In SAQ-C, because respondents had trouble clearly understanding the real meaning of power gradient in Chinese, the researchers changed the term. In CTT, this instrument theoretically requires a full-scale validation; if it passes, it would be called TPSC ver. 2.0. Yet in an item-oriented method like IRT, such an impact is minimized on the condition that local independence is observed, although we should report these changes to academia. Finally, IRT is not influenced by the survey taker; seeing such characteristics numerically is easy.^{28,29}

Conclusion

Time to shift gears to IRT in the field of safety culture

Thus far, only a few researchers in patient safety culture have utilized IRT, despite its superior applicability to various measurement scales. This might be due to many reasons, such as insufficient computing power for running such serious analyses in a short time, but to be honest, not many people really understand how it works. This article confined itself to the validation. The real-world use of IRT in safety culture surveys can be found in previously published articles.^{28,30,31}

References

- Reason J. Safety paradoxes and safety culture. *Injury Control and Safety Promotion*. 2000;7(1):3–14.
- Zhang H, Douglas A Wiegmann, Terry L von Thaden, et al. Safety culture: A concept in chaos? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. 2002;46(15):1404–1408.
- Guldenmund FW. The nature of safety culture: a review of theory and research. *Safety Science*. 2000;34(1–3):215–257.
- Jeong HJ, Su Mi Jung, Byung Joo Song. Combinational effects of clinical area and healthcare workers' job type on the safety culture in hospitals. *Biom Biostat Int J*. 2015;2(2):00024.
- Jeong HJ, Pham JC, Kim M, et al. Major cultural-compatibility complex: Considerations on cross-cultural dissemination of patient safety programmes. *BMJ Quality & Safety*. 2012;21(7):612–615.
- Fogg MJ. *Terraforming: engineering planetary environments*. 1st ed. Geoffrey A Landis, editor. PA: Society of Automotive Engineers; 1995.
- Mc Afee A, Brynjolfsson E. Big data: The management revolution. *Harvard Business Review*. 2012;90(10):61–67.

8. Etcheagaray JM, Thomas EJ. Comparing two safety culture surveys: safety attitudes questionnaire and hospital survey on patient safety. *BMJ Quality & Safety*. 2012;21(6):490–498.
9. Sexton JB, Helmreich RL, Neilands TB, et al. The Safety Attitudes Questionnaire: psychometric properties, benchmarking data, and emerging research. *BMC Health Service Research*. 2006;6(1):44.
10. Nguyen G, Gambashidze N, Ilyas SA, et al. Validation of the Safety Attitudes Questionnaire (short form 2006) in Italian in hospitals in the northeast of Italy. *BMC Health Services Research*. 2015;15(1):284.
11. Lee GS. Are healthcare workers trained to be impervious to stress? *BIOM BIOSTAT INT J*. 2015;2(2):28.
12. Jeong HJ, Lee WC. The pure and the overarching: An application of bifactor model to Safety Attitudes Questionnaire. *Biom Biostat Int J*. 2016;4(6):110.
13. Lee WC, Wung HY, Liao HH, et al. Hospital safety culture in Taiwan: a nationwide survey using Chinese version Safety Attitude Questionnaire. *BMC Health Services Research*. 2010;10(1):234.
14. Chalmers RP. mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*. 2012;48(6):1–29.
15. Muthén B, Asparouhov T. IRT studies of many groups: the alignment method. *Frontiers in Psychology*. 2014;5:978.
16. Jeong HJ, Lee WC, Liao HH. Importance of covariance in confirmatory factor analysis: Safety Attitudes Questionnaire–Chinese Version as an example. *Biom Biostat Int J*. 2017;6(2):00165.
17. Houts CR, Cai L. flexMIRT user’s manual version 3.5: Flexible multilevel multidimensional item analysis and test scoring. *Vector Psychometric Group*. 2017:1–240.
18. Cai L, Hansen M. Limited-information goodness-of-fit testing of hierarchical item factor models. *Br J Math Stat Psychol*. 2013;66(2):245–276.
19. Hansen M, Li Cai, Scott Monroe, et al. Limited-information goodness-of-fit testing of diagnostic classification item response theory models. *National Center for Research on Evaluation*. 2014.
20. Maydeu-Olivares A, Joe H. Limited-and full-information estimation and goodness-of-fit testing in 2 n contingency tables: a unified framework. *Journal of the American Statistical Association*. 2005;100(471):1009–1020.
21. Jeong HJ. Development of the Safety Attitudes Questionnaire–Korean Version (SAQ–K) and its novel analysis methods for safety managers. *Biom Biostat Int J*. 2015;2(1):00020.
22. Lee WC. Validation study of the Chinese Safety Attitudes Questionnaire in Taiwan. *Taiwan J Public Health*. 2008;27:6–15.
23. Embretson SE, Reise SP. *Item response theory*. New Jersey: Lawrence Erlbaum Associates; 2010.
24. Acock AC. *Discovering structural equation modeling using Stata*. Texas: Stata Press Books; 2013.
25. De Vellis RF. *Classical test theory*. *Medical Care*. 2006:S50–S59.
26. Fliege H, Becker J, Walter OB, et al. Development of a computer-adaptive test for depression (D–CAT). *Quality of Life Research*. 2005;14(10):2277.
27. Mead AD, Meade AW. Test construction using CTT and IRT with unrepresentative samples. *CTT and IRT*. p. 1–40.
28. Jeong HJ, Lee WC. Item response theory-based evaluation of psychometric properties of the Safety Attitudes Questionnaire—Korean Version (SAQ–K). *Biom Biostat Int J*. 2016;3(5):00079.
29. Thissen D, Steinberg L. Item response theory. *The Sage Handbook of Quantitative Methods in Psychology*. 2009:148–177.
30. Jeong HJ, Lee WC. Ignorance or negligence: Uncomfortable truth regarding misuse of confirmatory factor analysis. *Journal of Biometrics & Biostatistics*. 2016;7(3):298.
31. Liao HH, Lee WC, You YL, et al. A practical approach to develop a parsimonious survey questionnaire—Taiwanese Patient Safety Culture Survey as an example. *Biom Biostat Int J*. 2017;6(3):00169.