Shiraz E-Med J. 2021 December; 22(12):e110977. doi: 10.5812/semj.110977.
Published online 2021 August 16. **Research Article**

A Differential Item Functioning (DIF) Study of the Infection Control Standard Isolation Precaution Instrument Across Gender and Major Among Healthcare Workers

Amin Mousavi [1], Mohsen Momeni [2], Mina Danaei [3] and Mehrdad Askarian [4, 5, *]

[1]Department of Educational Psychology and Special Education, College of Education, University of Saskatchewan, Saskatoon, Canada
[2]Neuroscience Research Center, Institute of Neuropharmacology, Kerman University of Medical Sciences, Kerman, Iran
[3]Social Determinants of Health Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran
[4]Department of Community Medicine, School of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran
[5]Health Behavior Science Research Center, Shiraz University of Medical Sciences, Shiraz, Iran

[*]*Corresponding author*: Department of Community Medicine, School of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran. Email: askariam@sums.ac.ir

Received 2020 November 07; **Revised** 2021 June 09; **Accepted** 2021 June 17.

Abstract

Background: Measuring healthcare workers' (HCWs) knowledge, attitudes, and practices (KAPs) regarding isolation precaution is essential for infection control which needs a valid and reliable instrument.
Objectives: This study aimed to assess differential item functioning (DIF) across gender and major for the knowledge and practice items of the questionnaire, previously designed in Shiraz, Iran.
Methods: This cross-sectional survey was conducted on 1070 participants (males/females: 306/764; medical students/nurses: 466/624). The study instrument had three subscales with nine questions for each KAP subscale. The Mantel-Haenszel (MH) statistic was used. The DIF and differential test functioning (DTF) analyses were also performed in this study.
Results: There were very similar DIF outcomes for the knowledge and practice subscales, with one or two items indicating moderate DIF but comparable total scores across genders. Across majors, several items showed large DIF for both subscales. It was found that large DTF affects major for both subscales.
Conclusions: Our findings indicated large DIF and DTF levels of the questionnaire among medical students and nursing groups. More attention should be paid when developing the items. This study shows the importance of paying attention to valid evidence for instruments developed within the field of healthcare.

Keywords: Standard Isolation Precautions, Differential Item Functioning, Measurement Invariance, Psychometrics

1. Background

Healthcare-associated infections (HAIs) are a major health problem that can involve patients, visitors, or any healthcare workers (HCWs) in hospitals or other healthcare facilities and are considered a significant concern for the general public (1, 2). According to the Centers for Disease Control and Prevention (CDC), standard isolation precautions that have recommendations on topics including hand hygiene, personal protective equipment (PPE), respiratory hygiene/cough etiquette, patient placement, patient-care equipment, and instruments/devices, care for the environment, textiles and laundry, safe injection practices, infection control practices for special lumbar puncture procedures, and worker safety are the most critical and necessary infection control measures. If these measures are observed strictly by HCWs, the spread of microorganisms can be significantly reduced (3, 4).

Lack of knowledge, forgetfulness, shortage of time, limited resources, insufficient support by managers, and HCWs' socio-demographic characteristics, including gender, age, working site, experience, job category, and marital status, could affect HCWs' compliance with standard precautions (5). Measuring HCWs' knowledge, attitudes, and practices (KAPs) regarding infection control standard isolation precautions needs a valid and reliable instrument. A questionnaire is a common tool for data collection. The primary purpose of a questionnaire is to collect accurate (i.e., valid) and consistent (i.e., reliable) data (6). There is a paucity of credible instruments assessing HCWs' KAPs regarding infection control standard isolation precautions, especially in Iran. In an attempt to address this gap, Askarian et al. (7) developed a questionnaire for assessing HCWs' KAPs regarding infection control standard isolation precautions. Using a sample of 622 medical students, they found Cronbach's Alphas of 0.726, 0.765, and 0.782 for the knowledge, practice, and attitude tests, respectively. The content validity of the instrument was established by a group of experts comprising of infection control experts, medical experts, and a psychiatrist. It was also reviewed by experts from the Iranian national expert group of infection control specialists. In 2006, they also repeated a similar study on a sample of physicians, surgeons, surgical residents, and medical residents. Findings showed that, generally, median scores for knowledge and attitudes were moderate to high. Surgeons were the only group that revealed a moderate to strong ($r = 0.748$, P-value < 0.001) relationship between knowledge and attitudes, while for other medical groups, this relationship was weak. The mentioned study also showed that more than 80% of all medical practitioners had not received previous education on infection control standard isolation precautions, and more than 80% were willing to be trained (8). Finally, in 2007, Askarian et al. (9) administered the same questionnaires on a sample of nursing, assistant nursing, and midwifery practitioners and students, along with interviews. Their findings revealed that about 91% of participants needed additional infection control education, especially on standard isolation precautions. They found positive correlations between KAPs for nurses, assistant nurses, and midwifery instructors and students.

Based on previous studies, this instrument shows high promises for assessing HCWs' KAPs regarding infection control standard isolation precautions. KAPs regarding standard precautions among HCWs are limited, and previous studies have shown a variety of KAPs among different groups of HCWs (10-12). As part of the instrument validation, it is important to ensure the uniformity of assessment across different demographic groups. Such a study is known as measurement invariance (MI) (13). Technically, it is important to test whether the probability of responding to a specific item differs across different identifiable groups after controlling for the construct being measured (13). This is known as testing differential item functioning (DIF). Generally speaking, there are two types of DIF. The uniform DIF implies constant differences in the probability of responding to an item between groups along the continuum of respondent's knowledge or ability (14). Figure 1 shows an example of the uniform DIF.

Another type of DIF is known as the non-uniform or crossing DIF. A non-uniform DIF implies that the differences in the probability of responding to an item not only depend on the respondent's group membership but also depend on his/her knowledge or ability. In another word, there's an interaction between group membership and knowledge or ability. Figure 2 shows an example of the non-uniform DIF.

The best-case scenario is not observing the DIF. This means that the probability of correctly answering a question should only be dependent on the respondent's knowledge or ability, not on other external variables such as gender, social status, job status, or race. Ensuring MI or the absence of DIF at the item level will result in comparable scores across demographic groups (15). Lack of MI at the item level (or the presence of DIF) may result in imprecise group differences in the observed scores. MI evaluates the equivalency of a construct across groups or measurement occasions. MI is relevant to group comparisons (e.g., the analysis of variance (ANOVA)), comparing means across repeated measures (e.g., pretest-posttest designs), and comparing the relationships between constructs across groups. Measurement non-invariance (MNI), on the other hand, implies that the construct of interest is not comparable across groups or occasions. This means that, under lack of MI, simply using the observed scores for statistical analysis such as ANOVA would yield incorrect conclusions because it is unclear whether the observed differences between the groups are real differences or are due to different perceptions of individuals. Another occasion in which MNIs can generate incorrect conclusions is when assessing the effectiveness of interventions. Such studies usually involve pretest-posttest measurements, control-treatment groups, or both. Researchers should ensure that pretest-posttest scores or the scores of control-treatment groups yield similar meanings and psychometric properties. If pretest-posttest scores or the scores of control-treatment groups are non-invariant, then it will not be clear if an observed change is due to the intervention or a change in participant's perceptions of the construct under study. This emphasizes the importance of establishing MI before using the observed score for any statistical analysis (16).

2. Objectives

This study aimed to assess DIF across gender and major (medical students versus nursing practitioners) for the knowledge and practice items as, according to previous studies, these two tests have a similar way of scoring (i.e., correct/incorrect) and are more relevant for practical and

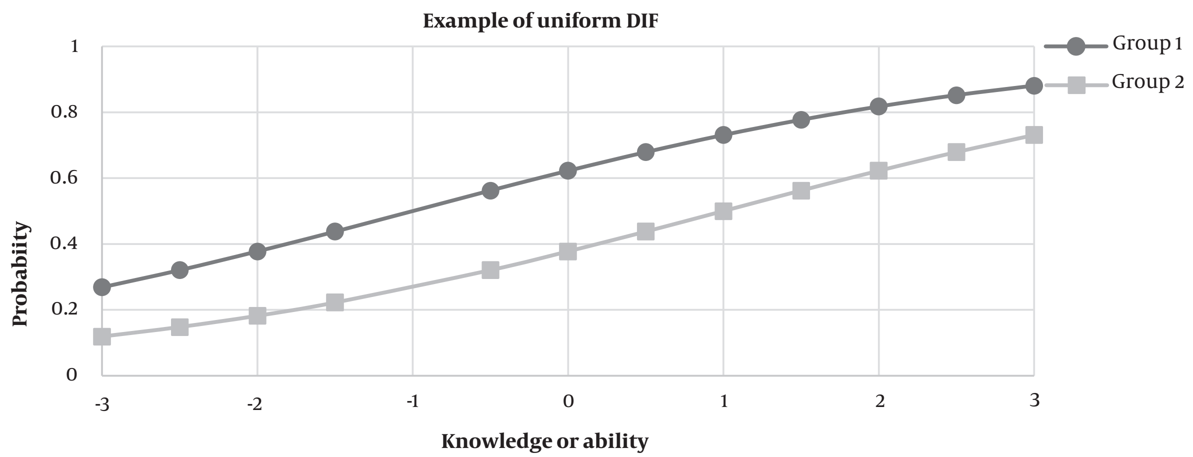

Figure 1. An example of the uniform DIF across two groups

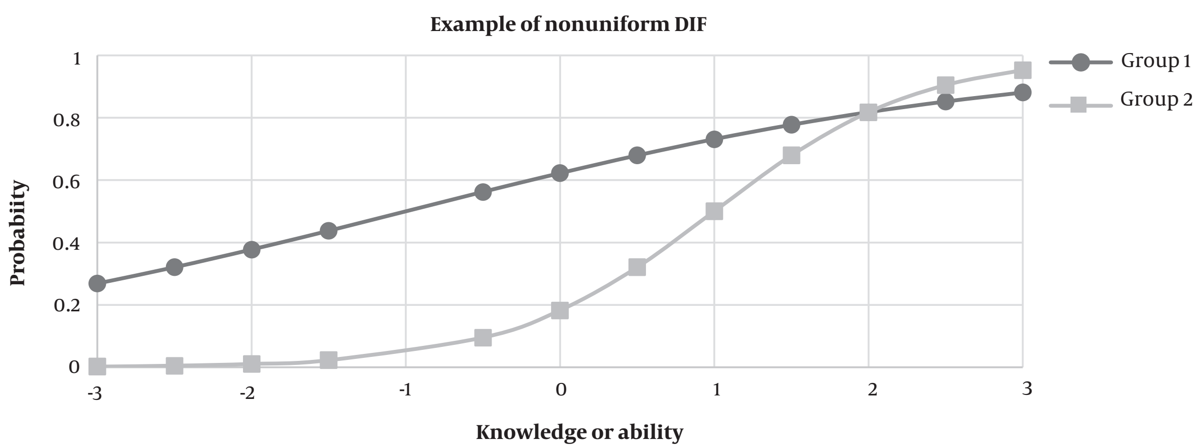

Figure 2. An example of the non-uniform DIF across two groups

educational purposes.

3. Methods

3.1. Participants and Data

A total of 1070 participants responded to the questionnaire. The sample consisted of 306 (28.6%) males and 764 (71.4%) females. There were 466 (41.7%) medical students and 624 (58.3%) nursing practitioners. The mean age of medical students was 27.32 ± 5.29, and that of nursing practitioners was 29.72 ± 6.21 years. The mean score for the knowledge subscale was 8.23 ± 1.2, while that for the practice subscale was 4.13 ± 2.95.

The ethical approval codes were obtained by the Ethical Committee of Shiraz University of Medical Sciences (code: IR.sums.med.rec.398.315). Verbal informed consent was obtained from all participants.

3.2. Instrument

The instrument had three subscales (including knowledge, practice, and attitude) with nine questions for each subscale. The questions related to the knowledge subscale had three possible answers (yes, no, and I don't know). All responses were scored as 1 for the correct answer and 0 for other answers. Therefore, the score for the knowledge test ranged between zero (no correct answer) and nine (all answers are correct). The questions assessing the practice subscale had five possible answers (always; often; sometimes; seldom; and never). A score of 1was allocated to 'always' and 0 to other answers. The total score ranged between zero and nine (all questions answered with 'always'). The goal of this subscale is to assess the extent to which HCWs practice the standards. In the present study, the reliability index using Cronbach's Alpha was 0.602 and 0.892 for the knowledge and practice subscales, respectively.

3.3. Data Analysis

The descriptive statistics and estimation of the reliability index were performed using IBM SPSS software, version 22. The DIF and DTF analyses were performed using the DIFAS 5.0 software (17)). To detect DIF, the Mantel-Haenszel chi-square (MH CHI) statistic (18, 19), the Mantel-Haenszel common log-odds ratio (MH LOR) (20), and the standardized Mantel-Haenszel log-odds ratio (SMH LORZ) were used. The MH CHI was distributed with chi-square with one degree of freedom, so the values greater than 3.84 were considered statistically significant at $\alpha = 0.05$. Positive values of the MH LOR indicated a DIF level in favor of the reference group, and negative values indicated a DIF level in favor of the focal groups. For the SMH LORZ, values higher than 2.0 or lower than -2.0 might denote the evidence for the presence of DIF. Furthermore, to classify the items showing DIF levels as small (A), moderate (B), and large (C), the ETS (Educational Testing Services) categorization scheme (21) was used. These categories were based on the differences calculated on the scale of item difficulty, known as Mantel-Haenszel delta difference (MH D-DIF). A statistically non-significant MH D-DIF of lower than 1.0 in an absolute value would categorize an item as A (i.e., small DIF). Items with MH D-DIF, which were significantly higher than 1.0 and an absolute value of 1.5 or higher would be categorized as C (i.e., large DIF). All other items would be categorized as B (i.e., medium DIF) (18). Given that all items were dichotomous, the variance of DIF across items (i.e., τ^2) was used to detect DTF. DTF compares a set of items across groups. In other words, it evaluates the comparability of total test scores (i.e., the observed scores) across groups. In this sense, it is an extension of DIF (22). As proposed by Penfield and Algina, values of $\tau^2 < 0.07$ indicate small, $0.07 \leq \tau^2 \leq 0.14$ indicate medium, and $\tau^2 > 0.14$ indicate large DTF (23). Two demographic characteristics were chosen for detecting DIF and DTF: Gender (the reference group = male, versus the focal group = female) and major (the reference group = medical students, versus the focal group = nursing practitioners).

4. Results

The results of DIF and DTF are reported in Tables 1 and 2. In assessing DIF and DTF across genders, the male category was set as the reference group, and regarding major, the medical category was set as the reference group.

A review of DIF values for the knowledge subscale (Table 1) shows that the majority of the items did not exhibit DIF across genders. Only item 2, "Washing hands before and after using gloves" (in favor of males), and item 8, "Bending needles before disposal" (in favor of females), showed moderate levels of DIF with significant MH CHI and a moderate level of DIF. At the scale level, a very small value of τ^2 was found, suggesting that the overall effects of the DIF items compensate each other, resulting in a comparable total score across genders. As for the medical group versus the nursing group contrast, two items showed significant MH CHIs with moderate DIF levels (i.e., item 6, "Washing hands with betadine after contact with patients during procedures and activities that are likely to generate splashes or sprays of blood and body fluids", in favor of the medical group and item 9, "Wearing a gown during procedures that are likely to generate splashes or sprays of blood and body fluids", in favor of the nursing group). Additionally, two items showed significant MH CHIs with large DIF levels. Item 1, "Washing hands before and after patient care", showed a large DIF level in favor of the medical group, indicating that members in this group had a significantly higher chance of correctly answering this question than those in the nursing group. In other words, just being in this group significantly increased their chance for a correct answer, and not necessarily their level of knowledge. On the contrary, item 8, "Bending needles before disposal", showed a large DIF level in favor of the nursing group. This means that being in this group significantly increased their chance of correctly answering this question. A value of $\tau^2 = 0.515$ indicated a large effect at the scale level, leading to incomparable total scores across majors.

As it is evident from the values in Table 2, all items of the practice subscale function similarly across genders, except for item 7, "Wearing a surgical mask to protect nose and mouth during procedures and activities that are likely to generate splashes or sprays of blood and body fluids", that showed a significant MH CHI with moderate DIF in favor of male respondents. Similar to the knowledge subscale, a very small value of τ^2 suggested a comparable total score across genders. The practice subscale showed more items with significant MH CHIs and moderate to large DIF levels across majors. Item 2 and item 5, "Wearing goggles to protect mucous membranes of the eyes during procedures

Table 1. The DIF and DTF Levels of the Knowledge Subscale[a]

Item	Gender				Major			
	MH CHI	MH LOR	SMH LOR Z	ETS	MH CHI	MH LOR	SMH LOR Z	ETS
1	3.19	0.71	1.90	A	20.03*	2.13	4.08[b]	C
2	4.93*	0.53	2.32[b]	B	0.23	0.12	0.59	A
3	0.11	0.03	0.04	A	0.14	-0.01	-0.02	A
4	0.11	-0.01	-0.02	A	0.01	0.32	0.45	A
5	0.25	-0.19	-0.66	A	0.44	-0.22	-0.80	A
6	0.12	0.12	0.46	A	4.90*	0.58	2.29[b]	B
7	0.33	0.26	0.74	A	0.63	0.32	0.95	A
8	9.65*	-0.69	-3.05	B	18.22*	-0.91	-4.16[b]	C
9	2.04	-0.49	-1.60	A	6.55*	-0.81	-2.68[b]	B
DTF	$\tau^2 = 0.003$				$\tau^2 = 0.515$			

Abbreviations: MH CHI, Mantel-Haenszel chi-square; MH LOR, Mantel-Haenszel common log-odds ratio; SMH LOR Z, standardized Mantel-Haenszel log-odds ratio; ETS, ETS categories for the magnitude of DIF.
[a] *, statistically significant at $\alpha = 0.05$.
[b] Item showing DIF.

Table 2. The DIF and DTF Levels of the Practice Subscale[a]

Item	Gender				Major			
	MH CHI	MH LOR	SMH LOR Z	ETS	MH CHI	MH LOR	SMH LOR Z	ETS
1	0.27	0.14	0.64	A	0.01	-0.01	-0.03	A
2	0.07	0.10	0.40	A	12.99*	-0.92	-3.6	B
3	0.73	-0.24	-0.97	A	2.22	0.43	1.61	A
4	0.39	-0.17	-0.75	A	22.77*	1.17	4.6	C
5	0.05	0.08	0.35	A	7.24*	-0.72	-2.82[b]	B
6	1.94	-0.27	-1.49	A	0.00	0.01	0.0	A
7	7.83*	0.61	2.88[b]	B	23.45*	1.19	4.83[b]	C
8	0.04	0.06	0.30	A	0.16	-0.09	-0.4	A
9	2.33	-0.44	-1.67	A	35.95*	-1.74	-5.80[b]	C
DTF	$\tau^2 = 0.035$				$\tau^2 = 0.754$			

Abbreviations: MH CHI, Mantel-Haenszel chi-square; MH LOR, Mantel-Haenszel common log-odds ratio; SMH LOR Z, standardized Mantel-Haenszel log-odds ratio; ETS, ETS categories for the magnitude of DIF.
[a] *, statistically significant at $\alpha = 0.05$.
[b] Item showing DIF.

that are likely to generate splashes or sprays of blood and body fluids", showed moderate levels of DIF, both in favor of the nursing group. Three items revealed large levels of DIF. Item 4, "Wearing gloves before touching mucous membranes and non-intact skin", and item 7 showed large DIF levels in favor of medical respondents, and item 9 showed large DIF in favor of nurse respondents. A value of $\tau^2 = 0.754$ indicated a large effect at the scale level, leading to incomparable total scores across majors.

5. Discussion

Our findings demonstrated very similar DIF outcomes for the knowledge and practice subscales with one or two items showing moderate DIF, but comparable total scores, across genders. Small or very small values of τ^2 implied that the observed DIF levels at the item level either were negligible or the items displaying DIF in favor of different groups nullified each other at the test level. Several factors can contribute to the DIF. For example, item 8, "Bending needles before disposal", showed moderate DIF in favor of female respondents and large DIF in favor of nurse respondents on the knowledge subscale. One possible reason could be the fact that nursing practitioners usually administer injections and, therefore, are responsible for the safe disposal of needles and many nurses are female. In contrast, item 7, "Wearing a surgical mask to protect nose and mouth during procedures and activities that are likely to generate splashes or sprays of blood and body fluids", showed moderate DIF in the practice subscale in favor of male respondents and large DIF in favor of the medical group. One potential explanation could be the fact that most splash generating procedures such as suturing, endotracheal suctioning, bronchoscopy, and invasive vascular procedures that need close distance to patients are administered by medical doctors in Iranian hospitals.

Item 9, "Wearing a gown during procedures that are likely to generate splashes or sprays of blood and body fluids", on the practice subscale showing large DIF in favor of nurses could be explained by the fact that these procedures (such as care for patients), with the possibility of an extensive splash of blood, body fluids, secretions, or excretions, are usually performed by nurses, so they know more about these types of procedures. Similarly, item 4, "Wearing gloves before touching mucous membranes and non-intact skin", on the practice subscale showing large DIF in favor of the medical group, could be because of the reason that most procedures need touching mucous membranes and non-intact skin which are performed by medical doctors in Iran.

This study aimed to examine DIF and DTF of the knowledge and practice subscales of the KAPs regarding the infection control standard isolation precautions instrument. Our findings indicated large DIF and DTF levels when the instrument was administered to the respondents from the medical and nursing groups. There could be several explanations for these findings, some of which will be discussed in the same section. This was an important step towards ensuring and enhancing the validity of this instrument. Our findings highlighted the fact that more attention should be paid when developing the items. The majority of the items, that exhibited large DIF in favor of a particular group, had contents or examples most familiar or relevant to that group. This adds unwanted irrelevant variance to the data, biasing the results of any statistical analyses. In terms of test fairness, we wanted our instruments to measure the construct of interest (e.g., the level of practicing infection control standard isolation precautions) regardless of respondents' demographic backgrounds, as these standards are set to be exercised by all HCWs. Fortunately, DIF and DTF levels were negligible across genders, but the revision of the items' contents when it comes to practice is needed. We hope that this study sets an example and shows the importance of paying attention to valid evidence for instruments developed within the field of healthcare.

Footnotes

Authors' Contribution: Study concept and design: MA and AM. Acquisition of data: MA. Analysis and interpretation of data: MA and AM. Drafting of the manuscript: MA, AM, MM, and MD. Critical revision of the manuscript for important intellectual content: MA, AM, MM, and MD.

Conflict of Interests: Authors declare no conflicts of interest.

Ethical Approval: The ethical approval codes were obtained by the Ethical Committee of Shiraz University of Medical Sciences (code: IR.sums.med.rec.398.315).

Funding/Support: The authors would like to thank the vice-chancellor of research, Shiraz University of Medical Sciences (no.: 20254), for their assistance in this project.

Informed Consent: Verbal informed consent was obtained from all participants.

References

1. Munoz-Price LS, Banach DB, Bearman G, Gould JM, Leekha S, Morgan DJ, et al. Isolation precautions for visitors. *Infect Control Hosp Epidemiol*. 2015;**36**(7):747–58. doi: 10.1017/ice.2015.67. [PubMed: 26017347].
2. Karki T, Plachouras D, Cassini A, Suetens C. Burden of healthcare-associated infections in European acute care hospitals. *Wien Med Wochenschr*. 2019;**169**(Suppl 1):3–5. doi: 10.1007/s10354-018-0679-2. [PubMed: 30680486].
3. CDC. *Center for disease prevention and control: Standard precautions for all patients*. 2020, [cited 27 Sep 2020]. Available from: https://www.cdc.gov/infectioncontrol/basics/standard-precautions.html.
4. Ed. SpihcIs. *standard precautions in health care*. 2016, [cited 26 Oct 2019]. Available from: https://www.who.int/csr/resources/publications/EPR_AM2_E7.pdf.
5. Haile TG, Engeda EH, Abdo AA. Compliance with standard precautions and associated factors among healthcare workers in Gondar University Comprehensive Specialized Hospital, Northwest Ethiopia. *J Environ Public Health*. 2017. doi: 10.1155/2017/2050635.
6. Taherdoost H. Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research. *SSRN Electron J*. 2016. doi: 10.2139/ssrn.3205040.

7. Askarian M, Honarvar B, Tabatabaee HR, Assadian O. Knowledge, practice and attitude towards standard isolation precautions in Iranian medical students. *J Hosp Infect*. 2004;**58**(4):292–6. doi: 10.1016/j.jhin.2004.07.004. [PubMed: 15564005].
8. Askarian M, McLaws ML, Meylan M. Knowledge, attitude, and practices related to standard precautions of surgeons and physicians in university-affiliated hospitals of Shiraz, Iran. *Int J Infect Dis*. 2007;**11**(3):213–9. doi: 10.1016/j.ijid.2006.01.006. [PubMed: 16837226]. [PubMed Central: PMC7110518].
9. Askarian M, Memish ZA, Khan AA. Knowledge, practice, and attitude among Iranian nurses, midwives, and students regarding standard isolation precautions. *Infect Control Hosp Epidemiol*. 2007;**28**(2):241–4. doi: 10.1086/510868. [PubMed: 17265414].
10. Acharya AS, Khandekar J, Sharma A, Tilak HR, Kataria A. Awareness and practices of standard precautions for infection control among nurses in a tertiary care hospital. *Nurs J India*. 2013;**104**(6):275–9. [PubMed: 24974532].
11. Atif ML, Brenet A, Hageaux S, Fave MH, Cochet C, Baticle E, et al. Awareness of standard precautions for 4439 healthcare professionals in 34 institutions in France. *Med Mal Infect*. 2013;**43**(1):10–6. doi: 10.1016/j.medmal.2012.11.004. [PubMed: 23290383].
12. Askarian M, Mirzaei K, Mundy LM, McLaws ML. Assessment of knowledge, attitudes, and practices regarding isolation precautions among Iranian healthcare workers. *Infect Control Hosp Epidemiol*. 2005;**26**(1):105–8. doi: 10.1086/502495. [PubMed: 15693417].
13. Mousavi A, Krishnan V. Measurement invariance of early development instrument (EDI) domain scores across gender and ESL status. *Alberta J Educ Res*. 2016;**62**(3):288–305.
14. Mousavi A, Shojaee M, Shahidi M, Cui Y, Kutcher S. Measurement invariance and psychometric analysis of Kutcher Adolescent Depression Scale across gender and marital status. *J Affect Disord*. 2019;**253**:394–401. doi: 10.1016/j.jad.2019.05.010. [PubMed: 31103804].
15. Borsboom D. When does measurement invariance matter? *Med Care*. 2006;**44**(11 Suppl 3):S176–81. doi: 10.1097/01.mlr.0000245143.08679.cc. [PubMed: 17060825].
16. Millsap RE. *Statistical approaches to measurement invariance*. 1st ed. Routledge; 2012.
17. Penfield RD. DIFAS: Differential Item Functioning Analysis System. *Appl Psychol Measur*. 2016;**29**(2):150–1. doi: 10.1177/0146621603260686.
18. Holland PW, Thayer DT. Differential item functioning and the Mantel-Haenszel procedure. *ETS Res Rep Series*. 1986;**1986**(2):i–24.
19. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst*. 1959;**22**(4):719–48. [PubMed: 13655060].
20. Camilli G, Shepard LA, Shepard L. *Methods for identifying biased test items*. **4**. Sage; 1994.
21. Zieky M. Practical questions in the use of DIF statistics in test development. In: Holland PW, Wainer H, editors. *Differential item functioning*. Lawrence Erlbaum Associates; 1993.
22. Camilli G, Penfield DA. Variance Estimation for Differential Test Functioning Based on Mantel-Haenszel Statistics. *J Educ Measur*. 1997;**34**(2):123–39. doi: 10.1111/j.1745-3984.1997.tb00510.x.
23. Penfield RD, Algina J. A Generalized DIF Effect Variance Estimator for Measuring Unsigned Differential Test Functioning in Mixed Format Tests. *J Educ Measur*. 2006;**43**(4):295–312. doi: 10.1111/j.1745-3984.2006.00018.x.