Development Paper : Substance Use and Abuse

Evaluating Middle School Student Smoking and Drinking Refusal Skills – Comparing Physical Education Teacher and School Health Nurse Ratings

Hideki Iwata*

*Faculty of Human Sciences, Kanazawa University
Kakuma, Kanazawa, Ishikawa 920-1192 Japan
iwata@ed.kanazawa-u.ac.jp
[Received May 18, 2010 ; Accepted February 10, 2011]

Purpose: To test the reliability of methods used to evaluate the refusal skills of middle school students pressured to smoke or drink by comparing the ratings given by physical education teachers and school health nurses.
Method: Forty second-year (8th grade) middle school students (20 boys and 20 girls) were each rated by 17 physical education teachers and 20 school health nurses as they were observed being pressured by peers to smoke or drink (based on both verbal and non-verbal items).
Results: Test-retest reliability showed a high reliability coefficient of 0.8 or higher for four out of five of the teachers for verbal exchange rating. Many non-verbal evaluations showed a reliability coefficient of 0.75 or lower. Two groups of physical education teachers and school health nurses were further divided into two groups based on work experience to evaluate reliability. As a result, the intra-class correlation coefficients (ICC) for verbal evaluations of refusal skills for both smoking and drinking were high when the more experienced physical education teachers were excluded. However, all four groups showed an inadequate reliability coefficient for non-verbal expressions. Additionally, it was observed that the experienced physical education teachers tended to be harsher in their rating of subjects while the experienced school health nurses tended to be more lenient.
Conclusion: The evaluation of the middle school student refusal skills was sufficiently reliable for verbal responses, but the evaluation of non-verbal responses did not exhibit sufficient reliability.

Keywords: refusal skills, prevention of smoking and drinking, reliability of third-party evaluators, middle school student

[School Health Vol.7, 1-7, 2011]

1. Introduction

In the 1970s, the validity of anti-smoking and drinking education in adolescents that included the use of psychosocial factors have been reported through research conducted mainly in the United States (Nozu and Tsunoda, 1992). Meanwhile, similar practical educational report has become disseminated throughout Japan that focused on coping with being asked by others to drink or smoke. It was found that most of evaluation methods evaluate the degree of student's increased desire to learn about drinking and smoking by analyzing children's writings and teachers' impressions (Iwata et al., 2006). Because of this, it has not been made sufficiently clear whether the acquisition of refusal skills is possible through these practices. Recent research coping with such invitations to engage in these behaviors involves self-evaluation through self-efficacy (Epstein et al., 2008; Trudeau et al., 2003), role playing, and other research that included a third-party evaluation of a subject actually refusing to engage in the behavior (Nichols et al., 2006; Wynn et al., 1997; Donaldson et al., 1994; Rhorbach et al., 1987).

In order to test the evaluation methods of middle school students' refusal skills, the authors analyzed the relationship between the role-playing observation method by third parties and the self-evaluation method by themselves used in another study (Iwata, 2009). In analyzing these third-party evaluations, it

was found that improvements needed to be made in the rating of non-verbal expressions. There are no reports in Japan of using third-party observations of children's refusal skills, but such evaluation indexes have been used in intervention evaluation research elsewhere (Wynn et al., 1997; Donaldson et al., 1994; Rohrbach et al., 1987; Katz et al., 1989; Sussman et al., 1993; Franzini et al., 1990; Sikkema et al., 1995).

The purpose of this study is to evaluate the intra-class correlation for 37 school staff (17 physical education teachers and 20 school health nurses) and to examine the difference from evaluations by their teaching experience to test the reliability of third party raters.

2. Methods

2.1. Refusal skills role-playing test

2.1.1. Student subjects

The subjects of this study were 40 middle school students from Ishikawa Prefecture (20 boys, 20 girls). All subjects were second-year (8^{th} grade) middle school students (ages ranged from 14-15 years).

2.1.2. Procedure

The procedure for filming the middle school student refusal skill role-playing exercise was as follows (Iwata, 2009):

The role-play exercise was conducted with each student individually in a classroom during the school lunch break or after class had ended. The role-play was conducted by two researchers (college students) who were adequately informed of the content of the study. One researcher was in charge of videotaping and managing the role-play (Researcher A). The second researcher served as the inviter (Researcher B, a college student of the same gender as the subject). The role-play was conducted as follows: 1) the subject enters the room, 2) Researcher A explains the content of the study to the student, 3) two scenarios are presented to the subject in which they are invited to drink or smoke, 4) the subject is given five minutes to think, 5) the role-play is then conducted (step 5 was videotaped). The subject and Researcher B sat two meters apart in chairs facing one another. The subject and Researcher B were filmed from a position four meters away.

The role-playing scenario was created in reference to "A plan for the junior high school health education" (Japanese Society of School Health, 2001). The subjects were pressured to smoke or drink by 1) explaining how it is a good stress reliever, 2) trivializing the risks of use, 3) applying peer pressure, or 4) norm setting. In the case of smoking, subjects were told that 1) "It calms you down when you are feeling anxious," 2) "It's not unhealthy if you only smoke a little," 3) "We're friends, right? Let's smoke together," 4) "We all understand that smoking is a bad thing, but that's what makes it cool. Why don't you just give it a try?"

2.2. Evaluating refusal skills by video observation

2.2.1. Raters

As raters, I selected physical education teachers and school health nurses who have the most opportunity on smoking and drinking prevention education among school staff. The videotapes of the forty subjects used in the study, along with a self-report questionnaire, were sent to 20 physical education teachers (10 males, 10 females) and 20 school health nurses (20 females) in Ishikawa Prefecture. When selecting the raters although I did not adopt random sampling, I considered not placing a disproportionate emphasis on some teaching experiences. Seventeen physical education teachers (8 males, 9 females) and 20 school health nurses (20 females) cooperated with the study, thereby resulting in a total of 37 raters. The survey was conducted during the period from January to February in 2007. The self-report questionnaire included questions on their age and length of work experience. The 37 raters were also asked to evaluate the refusal skills of the forty middle school students by watching the video.

2.2.2. Behavioral assessment procedure and measure

The raters were made to evaluate each subject's response to the researcher's four invitation prompts on a scale of 1 to 5. The scale included the following responses, 1) no response (1 point), 2) single refusal (2 points), 3) refusal accompanied by a reason (3 points), 4) refusal accompanied by a suitable suggestion (4 points), 5) refusal accompanied by multiple suggestions (5 points).

The criteria for evaluating non-verbal responses included 1) fluency, 2) voice tone, 3) posture, gesture, 4) eye-contact, facial expression, and 5) overall

assertiveness including the above. The 3-point scale included the following assessments, 1) unable, 2) not very able, 3) able.

2.3. Data Analysis

The physical education teachers and school health nurses were each separated into two groups based on length of work experience (under 15 years and over 16 years of experience). The intra-class correlation coefficients (ICC) were calculated and the reliability of the ratings for the four groups was tested (Tani, 1997). Additionally, the evaluation was similarly conducted with 5 teachers (only school health nurses) at greater than 3-week intervals and the reliability of those evaluations was also tested. The scores from the four groups was compared using the Kruskal-Wallis H test, and then a multiple comparison test was conducted using the Mann-Whitney U test on the significant differences found within the four groups.

3. Results

3.1. Testing reliability with the test-retest method

Table 1 displays the trial reliability coefficient for the five raters. Each verbal evaluation ICC for drinking and smoking were found to be 0.8 or higher, excluding rater E. However, 3 of the 5 raters had reliability below 0.75, which is generally seen as a good level of reliability (Fleiss, 1981). Rater B had a non-verbal smoking reliability of 0.58 and non-verbal drinking reliability of 0.66; Rater C had a non-verbal smoking reliability of 0.62 and non-verbal drinking reliability of 0.73. Rater E had a non-verbal smoking reliability of 0.81, which was high, but also had a non-verbal drinking reliability of 0.61, which was the lowest reliability found among the raters.

3.2. Comparing the different faculty groups' ratings

A few of the ICCs for each of the four groups of raters in the verbal category exceeded 0.75 for the five items, but nearly all were below that level (**Table 2**). Also, for the non-verbal evaluation ICCs, there were no values that exceeded 0.75 (0.33-0.57). The overall assertiveness category for the non-verbal evaluation also reported a low score (0.47-0.57).

One result of comparing the differences in ratings between the four groups was that the group of more experienced school health nurses tended to evaluate each item higher than the other groups. Conversely, the group of more experienced physical education teachers tended to evaluate each item lowers than other groups. This difference was not significant between the two less experienced rater groups.

4. Discussion

The purpose of the present study was to test the reliability of third-party observatory evaluations by comparing the evaluations of two groups of 20 school health nurses and 17 physical education teachers. Each of these two groups was subdivided into two groups based on length of work experience and reliability was tested after calculating the groups' ICC. Each evaluation score was compared in the four groups. While I was unable to get an adequate value

Table 1 Results of test-retest reliability coefficient scores (n = 40)

	Rater				
	A	B	C	D	E
Verbal evaluation					
Smoking	0.90	0.82	0.81	0.88	0.88
Drinking	0.92	0.82	0.88	0.92	0.70
Non-verbal evaluation					
Smoking	0.77	0.58	0.62	0.85	0.81
Drinking	0.78	0.66	0.73	0.84	0.61

Table 2 The average, standard deviation, median, and inter-group comparisons for each rating

	Health and physical education teacher								School health nurse										
	G1 (career≦15 years, N = 6)				G2 (career>16 years, N =11)				G3 (careerkih≦15 years, N = 7)				G4 (career>16 years, N = 13)						
	Mean	SD	Median	ICC	Mean	SD	Median	ICC	Mean	SD	Median	ICC	Mean	SD	Median	ICC	$\chi^2(3)$	p	Multiple Comparison¶
Verbal evaluation																			
Smoking																			
1 Stress reliever	2.87	0.87	2.91	0.61	2.86	0.97	2.84	0.52	2.98	0.86	3.04	0.65	3.06	0.88	3.11	0.71	14.94	0.00	G1,G2 < G4
2 Trivializing the harm	2.99	0.93	3.00	0.73	2.94	1.01	2.89	0.57	3.03	0.91	3.07	0.74	3.11	0.92	3.12	0.72	8.93	0.03	G2 < G4
3 Peer pressure	2.94	0.96	2.86	0.72	2.89	1.00	2.79	0.57	3.01	1.04	2.89	0.78	3.03	0.99	2.97	0.76	6.14	0.10	
4 Norm setting	2.72	1.01	2.59	0.76	2.61	1.03	2.50	0.59	2.67	1.04	2.56	0.75	2.77	1.01	2.69	0.74	7.50	0.06	
Drinking																			
1 Stress reliever	2.57	0.85	2.52	0.63	2.51	0.90	2.45	0.47	2.63	0.85	2.63	0.70	2.71	0.89	2.70	0.68	16.98	0.00	G2 < G4
2 Trivializing the harm	2.79	0.94	2.80	0.72	2.71	1.01	2.69	0.65	2.83	0.95	2.88	0.73	2.89	0.95	2.92	0.72	10.19	0.02	G2 < G4
3 Peer pressure	2.69	0.93	2.60	0.58	2.64	0.96	2.56	0.47	2.88	0.98	2.78	0.57	2.93	0.94	2.88	0.61	33.03	0.00	G2 < G3,G4 / G1 < G4
4 Norm setting	3.10	1.08	3.15	0.70	2.95	1.08	2.95	0.65	3.16	1.12	3.22	0.77	3.25	1.04	3.31	0.74	19.08	0.00	G2 < G3,G4
Non-verbal evaluation																			
Smoking																			
1 Fluency	2.06	0.71	2.08	0.49	2.04	0.73	2.05	0.48	2.10	0.75	2.14	0.44	2.15	0.65	2.18	0.45	6.29	0.10	
2 Voice tone	2.13	0.67	2.15	0.52	1.96	0.72	1.95	0.47	2.07	0.78	2.10	0.52	2.15	0.69	2.19	0.43	17.51	0.00	G2 < G1,G4
3 Posture	2.11	0.72	2.14	0.42	1.89	0.72	1.86	0.37	2.00	0.79	2.01	0.33	2.18	0.68	2.21	0.42	40.17	0.00	G2 < G1,G4
4 Facial expression	1.95	0.67	1.94	0.39	1.95	0.69	1.93	0.42	1.94	0.74	1.92	0.35	2.12	0.68	2.15	0.35	22.30	0.00	G1,G2,G3 < G4
5 Overall assertiveness	2.05	0.64	2.06	0.50	1.97	0.70	1.96	0.49	2.06	0.73	2.08	0.48	2.20	0.64	2.23	0.47	28.55	0.00	G1,G2,G3 < G4
Drinking																			
1 Fluency	2.05	0.73	2.07	0.55	1.96	0.72	1.95	0.54	2.05	0.75	2.06	0.47	2.11	0.69	2.14	0.52	10.79	0.01	G2 < G4
2 Voice tone	2.03	0.69	2.03	0.53	1.87	0.74	1.83	0.52	2.00	0.76	2.00	0.48	2.07	0.73	2.10	0.51	19.79	0.00	G2 < G1,G4
3 Posture	2.04	0.70	2.05	0.45	1.85	0.72	1.81	0.43	1.95	0.80	1.92	0.47	2.12	0.69	2.15	0.46	34.76	0.00	G2 < G1,G4 / G3 < G4
4 Facial expression	1.95	0.70	1.93	0.50	1.91	0.68	1.88	0.41	1.86	0.74	1.83	0.37	2.08	0.69	2.10	0.44	22.75	0.00	G2,G3 < G4
5 Overall assertiveness	1.99	0.66	1.98	0.57	1.92	0.71	1.90	0.50	2.03	0.73	2.04	0.50	2.15	0.66	2.18	0.54	27.94	0.00	G1,G2,G3 < G4

¶: Mann-Whitney U test

for the non-verbal ratings, I was able to confirm a fixed reliability for the verbal evaluation. The reason for this is because it was observed to be easy to assign an evaluation standard based on the student-subject's comments. In addition, in order to improve reliability, it may be effective to adopt a multidimensional evaluation. Nichols et al. (2006) performed evaluation by the reaction of whether for some verbal strategies such as "Simple no", "Declarative statements as *I do not want to smoke*", and "Other reasons" can be seen (yes/no). Thus, the multidirectional evaluation can be expected to raise reliability of verbal evaluations that are used in this study. In the future, I expect to use the verbal evaluation as an intervention program evaluation index, such as was developed previously by Rohrbach et al. (1987).

Unfortunately, I was unable to get an adequate ICC for the non-verbal evaluation similar to what was previously reported (Iwata, 2009). The results of testing the ICC reliability for the physical education teachers and the school health nurses revealed that the non-verbal evaluation ICC was generally below the 0.75 standard that had been previously set (Fleiss, 1981). Moreover, as a new index for non-verbal evaluation, as it was used in previous studies (Nichols et al., 2006; Sussman et al., 1993; Franzini et al., 1990; Sikkema et al., 1995), I included general assertiveness criterion in this study. However reliability coefficient is not sufficiently high. It will be necessary to raise reliability by adopting multidimensional evaluation by "able ― not able" in each item also in non-verbal evaluation. Furthermore, the items to be evaluated will have to be narrowed down to appropriate numbers.

It has been inferred that one reason adequate non-verbal evaluation reliability was not obtained was due to the apparent necessity of providing training before conducting the video-based evaluations. Many researchers reported that the raters participate in evaluation training before conducting the observation evaluation (Nichols et al., 2006; Wynn et al., 1997; Donaldson et al., 1994; Rohrbach et al., 1987). Rohrbach et al. (1987) and Sussman et al. (1993) reported that they conducted evaluation training with the raters until the evaluation results increased in consistency while watching the video or actually being present in the classroom and watching the students who role-play as they interacted with the researcher. In the present study, I attempted to simplify the evaluation process by attempting to make the evaluation process easier. Unfortunately, this was not adequate to facilitate consistency in this process.

It is also believed that there may be problems in making reasonable comparisons with non-verbal expressions that are included in the students' refusal skills. In other words, one point of this study revealed whether there was a reasonable relationship between the evaluation index what is typically used as "non-verbal expressions" when middle school students are pressured to smoke or drink and such index used as non-verbal expressions" in this study. For example, Epstein et al., (1997) identified five strategies for handling such refusals that were directly taken from adolescents; these strategies were developed into evaluation criteria. These five strategies include 1) saying "no or, I am okay," 2) saying "I am okay right now," 3) changing the subject, 4) saying "I don't want to do that," or 5) making an excuse and walking away. Additionally, Nichols et al. (2006) tested the differences in strategies between two different scenarios, either being pressured to smoke or to engage in shoplifting. The refusal strategy used for smoking in this case was to use the "broken record" strategy of repeatedly saying "no" and other strategies were not widely used in this case. In the present study, I took up strategies that were referenced positively in previous studies, such as way of speaking, facial expression, posture, and gestures as having hypothetical prospects. However, it is natural to believe that there is an optimal refusal strategy for a given child. There is also the possibility that differences in who is pressuring the child to do what behavior will result in different refusal strategies. It will be necessary to more closely examine the non-verbal refusal skills that should be evaluated while I continue to identify easy refusal method strategies for adolescents who are pressured to smoke or drink in Japan.

In addition, Rohrbach et al. (1987) and Donaldson et al. (1994) evaluated subjects on the basis of whether they incorporated refusal strategies that they had acquired in an intervention program, such as "shaking no and saying nothing," or "staring at the person pressuring smoking or drinking." It will likely also be necessary to specify which of these non-verbal expressions are easy to use and then test such expressions using an evaluation method.

As stated earlier, in the present study, non-verbal criteria were used which included 1) fluency, 2) voice tone, 3) posture, gestures 4) eye-contact,

facial expression, and also a "general assertiveness" criterion which was included as it had been used in previous studies (Nichols et al., 2006; Sussman et al., 1993; Franzini et al., 1990; Sikkema et al., 1995). These factors were calculated for each of the four groups and compared. The result of this comparison showed that experienced physical education teachers rated the subjects more harshly, while experienced school health nurses rated the subjects more leniently. Further, there was no significant difference found between the two groups that were less experienced. Murakami et al. (2001) reported that in an evaluation of patient abilities that less experienced health care providers tended to score the subjects more leniently. The present results are consistent with Murakami et al. (2001), in those physical education teachers who had greater experience tended to evaluate the student-subjects more harshly. However, as stated previously, there was inadequate reliability in this instance. Also, the experienced school health nurses in this study indicate an opposite tendency to those reported previously by Murakami et al. (2001). It is also possible that there are different factors influencing the school health nurses evaluations in this study, since they are not being asked to evaluate a role-playing scenario that is directly related to their daily work as a medical worker needs to estimate a patient's capability. It is also possible that the two rater groups who had less experience were also younger and this may have influenced their perception of the middle school students' non-verbal expressions.

Among all four groups, the non-verbal expression rating for 4) "eye-contact, facial expression" corresponds to "not very able" below 2.00. This result may indicate that non-verbal expressions such as staring sharply at the researcher or changing facial expressions were not used relative to other non-verbal expressions. The raters may have also tended to harshly evaluate these types of non-verbal strategies. In any event, a more detailed study of these factors is in order.

Finally, I discuss the limits of this research. It is possible that, in the present study, the subjects did not fully express themselves non-verbally, thus making an evaluation difficult. One such factor that could have led to this result is the fact that the subjects were made to sit in a chair as the role-play was conducted. It is predicted that if the subject were standing, they might engage in other refusal strategies, such as approaching, backing away, or walking away. Some previous studies indicate that the subjects did engage in strategies such as walking away from the role-playing scene (Nichols et al., 2006; Rohrbach et al., 1987). It will therefore also be necessary to experiment by differing the setting of the role-play exercise in future studies.

Acknowledgements

The author wishes to thank the teachers in Ishikawa Prefecture who cooperated with this study. The author is also grateful to Professor Yuji Nozu (Graduate School of Comprehensive Human Sciences, University of Tsukuba) for his guidance and advice.

This work was supported in part by a Grant-in-Aid for Young Scientists (B) (No. 18700543) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT).

References

Donaldson, S.I., Graham, J.W., Hansen, W.B. (1994) Testing the generalizability of intervening mechanism theories: understanding the effects of adolescent drug use prevention interventions. Journal of Behavioral Medicine 17(2):195-216.

Epstein, J.A., Botvin, G.J. (2008) Media resistance skills and drug skill refusal techniques: What is their relationship with alcohol use among inner-city adolescents? Addictives Behaviors 33(4):528-537.

Epstein, J.A., Botvin, G.J., Diaz, T., Baker, E., Botvin, E.M. (1997) Reliability of social and personal competence measures for adolescents. Psychological Reports 81(2):449-450.

Fleiss, L. L. (1981) Statistical method for rates and proportions. Toronto : John Wiley & Sons.

Franzini, L.R., Sideman, L.M., Dexter, K.E.,et al. (1990) Promoting AIDS risk reduction via behavioral training. AIDS Education and Prevention 2(4):313-321.

Iwata, H. (2009) Predictors to evaluate high school students' skill at refusing tobacco and alcohol: Relationships among self-assessment and role-playing tests. The Bulletin of Japanese Curriculum Research and Development 32 (3):61-70 (in Japanese with English Summary).

Iwata, H., Nozu, Y., Watanabe, M., (2006) Role-playing in school health education: A review for Japan and the other countries. Japanese Journal of school Health 47 (6):510-524(in Japanese with English Summary).

Japanese Society of School Health. (2001) A plan for the junior high school health education (In Japanese).

Katz, R.C., Robisch, C.M., Telch, M.J.(1989) Acquisition of smoking refusal skills in junior high school students. Addictives Behaviors 14(2):201-204.

Murakami, J., Takenaka, H., Horikoshi, A., et al. (2001) Problems in the evaluation of medical interviewing skills with objective structured clinical examinations: How can reasonable objectivity be ensured? Medical Education 32(4):231-237 (In Japanese).

Nichols, T.R., Graber, J.A., Brooks-Gunn, J., Botvin, G.J. (2006) Ways to say no: refusal skill strategies among urban adolescents. American Journal of Health Behaviors 30(3):227-236.

Nozu, Y., Tsunoda, H. (1992) A review of studies on school-based smoking prevention programs. Japanese Journal of Public Health 39 (6):307-318 (in Japanese with English Summary).

Rohrbach, L.A., Graham, J.W., Hansen, W.B., et al. (1987)

Evaluation of resistance skills training using multitrait-multimethod role play skill assessments. Health Education Research 2: 401 - 407.

Sussman, S., Stacy, A.W., Dent, C.W., Burciaga, C., Burton, D., Flay, BR. (1993) Refusal assertion versus conversational skill role-play competence: relevance to prevention of tobacco use. Statistics in Medicine 12(3-4):365-376.

Sikkema, K.J., Winett, R.A., Lombard, D.N. (1995) Development and evaluation of an HIV-risk reduction program for female college students. AIDS Education and Prevention 7(2):145-159.

Tani H.,(1997) Reliability of evaluation. Rigakuryoho Kagaku 12(3):113-120 (in Japanese with English Summary).

Trudeau, L., Spoth, R., Lillehoj, C., Redmond, C., Wickrama, K.A. (2003) Effects of a preventive intervention on adolescent substance use initiation, expectancies, and refusal intentions. Prevention Science 4(2):109-122.

Wynn, S.R., Schulenberg, J., Kloska, D.D., Laetz, V.B. (1997) The mediating influence of refusal skills in preventing adolescent alcohol misuse. Journal of School Health 67(9):390-395.

Name:
Hideki Iwata

Affiliation:
Faculty of Human Sciences, Kanazawa University

Address:
Kakuma, Kanazawa, Ishikawa 920-1192 Japan

Brief Biographical History:
2001- Assistant Professor, Faculty of Education, Kanazawa University
2007-Associate Professor, Faculty of Human Sciences, Kanazawa University
2010- Professor, Faculty of Human Sciences, Kanazawa University

Main Works:
• "Predictors to evaluate high school students' skill at refusing Tobacco and alcohol: Relationships among self-assessment and role-playing tests." The Bulletin of Japanese Curriculum Research and Development, Vol.32,61-70,Dec. (2009).

Membership in Learned Societies:
• Japanese Association of School Heath
• Japan Society of Physical Education, Health and Sport Sciences
• Japan Curriculum Research and Development Association