

Reliability and validity of the Japan Ijime Scale and estimated prevalence of bullying among fourth through ninth graders: A large-scale school-based survey

Yuko Osuka, BA ^{1,2} Tomoko Nishimura, PhD,^{1,2,3} Manabu Wakuta, PhD,^{2,4} Nori Takei, MD, PhD^{1,3,5} and Kenji J. Tsuchiya, MD, PhD^{1,2,3*}

Aim: The present study aimed at developing a novel scale, the Japan Ijime Scale (JaIS), to measure bullying in Japan with substantial reliability and validity, with which we estimated the prevalence of bullying among children and adolescents of school age.

Methods: The JaIS is a self-report questionnaire and consists of three parts: subscales measuring victimization and witnessing, and an item measuring perpetration. To test the reliability and validity of the two subscales, the authors analyzed responses to the JaIS from 2334 school students (Grades 4–9) in six elementary and three junior high schools in a middle-sized industrial city in central Japan, using exploratory factor analysis, item response theory, and examination of the external validity of the items. The prevalence of bullying victimization, witnessing, and perpetration was estimated.

Results: Item response theory models revealed that both the Victimization and Witness subscales have sufficient discrimination power and measurement precision, and the external validity of each scale has been confirmed. Using the JaIS, we found that 35.8% of students had been victims of bullying every 2–3 months (27.6% were solely victims and 8.3% were bully/victims), 32.8% had witnessed some type of bullying act, and 11.8% had perpetrated some type of bullying (3.5% as perpetrators, and 8.3% as bully/victims).

Conclusion: The JaIS is a reliable and valid measure. Using this scale, we found a high prevalence of bullying victimization in Japanese schools.

Keywords: aggression, bullying victimization, Japan, psychometrics, school bullying.

<http://onlinelibrary.wiley.com/doi/10.1111/pcn.12864/full>

Bullying among children and adolescents is a worldwide phenomenon, often associated with adverse effects on students' mental health,^{1,2} school absenteeism,³ and poor academic achievement.^{4,5} In addition to short-term effects, bullying has numerous undesirable long-term behavioral consequences for victims, such as depression, anxiety, self-mutilation,^{6–10} suicide, or suicide attempts.^{11,12}

Even though the term *bullying* tends to be used in the media as a blanket term that encompasses aggressive or negative acts among individuals, social-scientific studies over the past few decades have focused on how *bullying* should be defined. The definition by Olweus⁶ is one of the most widely accepted. It consists of three components – 'intentionality to harm (an)other person(s),' 'repetitiveness,' and 'power imbalance' – all observed during 'the past couple of months.'⁶

In Japan, the government's 'Act for the Measures to Prevent Bullying'¹³ defines *school bullying* (*ijime* in Japanese) as 'an act by a student, or students, toward another student that inflicts some physical or psychological consequence causing the receiving child mental or physical suffering.' Other studies from Japan have defined *bullying* in different ways and have incorporated terms such as 'damage to victim'¹⁴ and 'a plurality of perpetrators.'^{15–17}

Using Olweus's definition,⁶ Solberg and Olweus¹⁸ reported a 10.1% prevalence of bullying victimization among fourth to ninth graders in Norway. Morita¹⁹ conducted a comparative study in Japan using Olweus's definition and reported prevalence of 8.5% during one semester among fifth to ninth graders. In turn, the National Institute for Educational Policy Research in Japan reported the prevalence of bullying victims during one semester among first to sixth graders as 45.8% for boys and 51.5% for girls.¹⁶ The higher prevalence reported in this study compared to the previous ones is likely due to the different definition of *bullying*, possibly leading to an overestimation. Other researchers in Japan have reported varying estimates based on varying definitions. Taki¹⁵ reported the prevalence of bullying by plural perpetrators, lasting for a couple of weeks, among fifth to ninth graders as 4.0%. Murayama *et al.*²⁰ reported the prevalence of bullying victimization among fourth and ninth graders as ranging from 5.5% (girls, sixth grade) to 16.0% (boys, fifth grade). Other Japanese studies have also reported the prevalence of 7.4%²¹ and from 10.1% to 14.5%,²² respectively. It is significant to note that the definition of *bullying* was not presented to the participants in these studies, in contrast to the previous studies. Furthermore, the participants were asked to sign their name in one of these studies.²⁰ Shimoda found that the reported prevalence of bullying victimization decreased significantly

¹ United Graduate School of Child Development, Hamamatsu University School of Medicine, Shizuoka, Japan

² Institute of Child Developmental Science Research, Shizuoka, Japan

³ Research Centre for Child Mental Development, Hamamatsu University School of Medicine, Shizuoka, Japan

⁴ United Graduate School of Child Development, Osaka University, Osaka, Japan

⁵ Institute of Psychiatry, King's College London, London, UK

* Correspondence: Email: tsuchiya@hama-med.ac.jp

for questionnaires that required students to sign their names, compared to unregistered (or anonymous) questionnaires.²³ The decrease is probably because students who have been bullied are afraid that the bullying will worsen if other students learn who reported the victimization.

Various scales have been used to estimate the prevalence of bullying. The revised Olweus Bully/Victim Questionnaire (OBVQ)²⁴ is one of the most frequently used scales to measure school bullying. This scale was intended to incorporate Olweus's definition, with its reliability and validity established.^{25,26} In Japan, Morita¹⁹ translated and used the OBVQ to measure bullying victimization, perpetration, and witnessing, but did not test the reliability or validity of the translated scale. Additionally, no other existing measures available in Japan have been tested for the psychometric property of the scale.

As for the psychometric properties of the established scales, Olweus²⁷ studied the factor structure of bullying and reported that bullying victimization is unidimensional. Olweus also indicated that *cyberbullying* (a relatively new aspect of bullying that is treated separately in the literature²⁸) can be understood according to his definition.⁶ This was further supported by a Brazilian study²⁶ and was suggested by Japanese studies,^{29–31} although these Japanese studies have not conducted a factor analysis to support unidimensionality.

Another concern inherent to bullying research, particularly in Japanese studies, is that the witnessing of bullying (i.e., as distinct from victimization and perpetration) has not been thoroughly considered. Victims of bullying are not always ready to ask for help by themselves at the time of the victimization.³² When facing the event, witnesses can help the victim and ask for someone to help.³³ As such, the potentially helpful role of the witness has been suggested as being important.^{34–37} Although two Japanese studies showed the prevalence of witnessing bully events (39% to 48%),^{34,38} they did not use any validated measurement tool.

The first aim of this study was to address the shortcomings of the existing literature, and to develop a reliable and valid scale to measure bullying prevalence – the Japan Ijime Scale (JaIS) – based on the definition of *bullying* by Olweus.⁶ We adopted the following definition of *bullying* in the Japanese context: *Bullying (ijime)* is an act with the 'intention to harm (an)other person(s)' where there is an 'imbalance in power between the victim and perpetrator(s)' and with a 'repetitive nature.' The second aim of this study was to estimate separately the prevalence of bullying victimization, witnessing, and perpetration in Japan, and to compare these results with those reported in the literature.

Methods

Participants

The study participants comprised 1273 students (Grades 4–6) in six elementary schools and 1061 students (Grades 7–9) in three junior high schools, located in a middle-sized industrial city with a population of approximately 170 000 inhabitants. This research was carried out at the request of the city's Board of Education. We made no selection of study participants; rather, we included all the students of all the school classes assigned to participate in the survey by the Board and randomly selected the classes between the fourth and ninth grades among the available schools. Out of a total of 2384 participants, 97.9% (2334) had available data.

Japan Ijime Scale

To measure bullying within a Japanese context, we developed the JaIS, a self-report questionnaire consisting of three parts: a Victimization subscale, a Witnessing subscale, and a Perpetration item. The Victimization and Witnessing subscales include nine questions each, one for each type of bullying identified in the revised OBVQ²⁴: physical bullying, verbal bullying, social exclusion/isolation, having money or other things taken or damaged, lies and false rumors, being threatened or forced to do things, racial bullying, sexual bullying, and cyberbullying. We added just one item concerning perpetration

(i.e., whether the student had ever perpetrated bullying) because we were prohibited by the local Board of Education from asking students more than one question about perpetration because of ethical considerations. The stem question for the Victimization subscale, Witnessing subscale, and Perpetration item was: 'Have you experienced any of the following events in the last 2 or 3 months?'

The response options for the Victimization subscale and for the Perpetration item were as follows: *Nothing in the past 2 or 3 months*, *Only once or twice in the past 2 or 3 months*, *Two or 3 times a month in the past 2 or 3 months*, *About once a week in the past 2 or 3 months*, and *Several times a week in the past 2 or 3 months*, coded on a 5-point scale from 0 to 4; the statistically supported cut-off point was investigated in the later analyses.

The response alternatives for the Witnessing subscale were: *I have never seen or heard anything* or *I have seen or heard something*, coded as 0 or 1, respectively. We did not aim at estimating the frequency of victimization by measuring witness frequency, as the previous studies on the prevalence of witnessing by Yonezato³⁴ or Matsushita³⁸ did not measure the frequency of witnessing. Therefore, a binary response for the Witnessing subscale was sufficient for our current aims. The first section of the questionnaire described personal data protection to explain to the students why we had employed an unregistered form. To ensure that all students had an equivalent comprehension of the concept of *bullying*, we provided explanatory notes in the questionnaire and also in the teachers' instruction notes to delineate the difference between 'playful teasing' and 'bullying.' The instructions were as follows:

We are asking if you are receiving '*ijime* or what might be *ijime*' from other students. But we do not call it '*ijime* or what might be *ijime*' when teasing is done in a friendly or playful way, as in *asobi* [play] or *jareai* [being playful] and so on. Also, it is not '*ijime* or what might be *ijime*' when two students have the same strength or equivalent position.

Demographic information

Students' demographic data were collected concerning grade, sex, nationality, household income, and language spoken at home to investigate the difference in prevalence.

Ethical issues

The study protocol was approved by the Hamamatsu University School of Medicine and the University Hospital Ethics Committee. The study's purpose, significance, and methodology were explained to students. They were also informed that they would accrue no disadvantage by not participating in the study. Oral assent was obtained from the students and written consent was obtained from their parents or guardians.

Statistical analyses

Analyses were conducted separately for the JaIS Victimization and Witnessing subscales, respectively, using the following procedure.

Exploratory factor analysis

First, exploratory factor analysis (EFA) was performed on the two subscales. In EFA, all items were treated as categorical. Weighted least squares, which are robust to non-normality, were used for estimation. We assumed that either the single- or the two-factor solution would fit, as Olweus²⁷ proposed a single factor of 'bullying' while Kubiszewski and colleagues²⁸ proposed two factors of 'cyberbullying' and 'traditional bullying' (the latter including verbal, physical, and indirect or relational bullying²⁷).

In EFA, we examined the scree plot, eigenvalues, and magnitude of factor loadings. Model fit was evaluated based on the following indices: root-mean-square error of approximation (RMSEA), comparative fit index (CFI), Tucker–Lewis index (TLI), and the standardized root-mean residual (SRMR). For the RMSEA, a value smaller than 0.05

indicates a good fit, while values of CFI and TLI above 0.90 are considered adequate.³⁹

Item response theory

Item response theory (IRT) models are mathematical equations describing the association between a respondent's underlying level on a latent trait and the probability of a particular item response, using a nonlinear monotonic function.⁴⁰ After confirming the unidimensionality of the scales, IRT was used to determine the validity and reliability of the subscales using the total sample.^{40,41}

The two-parameter logistic model

The Witnessing subscale was analyzed with the two-parameter logistic model⁴² of IRT because the response category was dichotomous. In the two-parameter logistic model, slope (item discrimination) parameters and location (item difficulty) parameters were estimated for each dichotomized item. An item with a higher slope parameter value meant that the item had a high ability to differentiate students who had witnessed someone being bullied from those who had not. The location (difficulty) parameter indicates the trait level (i.e., level of witnessing) required to have a 50% chance of endorsing an item. Accordingly, an item with a higher location parameter value meant that students with higher latent trait levels (i.e., levels of witness) had endorsed the item.

The graded response model

The Victimization subscale was analyzed with the graded response IRT model⁴³ because the response category was more than two and ordinal. The graded response model is an extension of the two-parameter logistic model,⁴² in which each item is described by a slope parameter and between-category threshold parameters (one less than the number of response categories).⁴⁰ A high slope parameter value suggests that an item has a high ability to differentiate between students who have been bullied and those who have not. Threshold parameters represent the trait level (i.e., level of victimization) necessary to respond above the threshold with a 0.50 level of probability.

Breivik and Olweus²⁵ reported that three categories with higher frequencies of victimization were collapsed due to the low number of responses and the analyses comprised three response options (i.e., *Not bullied*, *Bullied once or twice*, or *Bullied more often*). The intersections of item characteristic curves (ICC) were examined to verify the number of categories.⁴⁰ ICC represent category probabilities of endorsing an item as a function of the latent trait. Each category probability shows a peak value at a given latent trait level. The categories *Around once a week* and *Several times a week* were collapsed into one category because the response probability of the former was below the probability of other categories.⁴⁰ We investigated the feasibility of amalgamating categories that had response probabilities below those of other categories into other categories that had higher response probabilities.

The reliability of the two subscales was examined using the item information functions and test information functions obtained in the IRT models. In IRT models, measurement precision can potentially differ for people with different trait levels. Unlike classical test theory in which measurement precision is typically represented by a single number (such as Cronbach's alpha), in IRT there are as many standard errors of measurement as there are unique trait estimates.⁴⁰

The item information curve is plotted to represent relative information as a function of trait level, called *theta*, reflecting the level of experience of bullying victimization or witnessing. The test information curve represents the relative precision of the scale across different levels of the trait continuum, and the height of this curve is proportional to the standard error of measurement,⁴⁴ with its highest point representing the highest precision, namely, the highest reliability, of the scale.

External validity

To test the external validity of the Victimization and Witnessing subscales, we used the abbreviated version of the Depression Self-Rating Scale for Children (DSRS-C).⁴⁵ The scores of the two subscales were summed up, respectively. Total scores for the Victimization subscale ranged from 0 to 36 points (0–4 points × nine items), and those of the Witnessing subscale ranged from 0 to 9 points (0–1 points × nine items). The DSRS-C abbreviated version is composed of nine items; each item is rated on a 3-point Likert-type scale. The total scores of the DSRS-C ranged from 0 to 36; a higher total score indicates a more severe depressive mood in students. The total scores of the DSRS-C were standardized in the analysis. We examined the association between Victimization subscale and DSRS-C scores using a negative binomial regression analysis, considering the distribution, and adjusting for students' sexes and grades. We also examined the association between Witnessing subscale and DSRS-C scores in the adjusted negative binomial regression analysis.

We tested the external validity of the Perpetration item using the score for Conduct Problem, a subscale of the Strengths and Difficulties Questionnaire (SDQ).⁴⁶ The SDQ consists of five subscales (of five items each) with each item rated on a 3-point scale. The Conduct Problem subscale ranges from 0 to 15; a higher score indicates that a student has a more severe conduct problem. The scores of the SDQ Conduct Problem subscale were standardized in the analysis. Scores for Perpetration ranged from 0 to 4 points. The association between Perpetration and Conduct Problem scores was also examined by adjusted negative binomial regression analysis.

Results

Sample characteristics

Data were obtained from 2334 students (51.2% female). Their demographic characteristics are shown in Table 1.

Exploratory factor analysis

Victimization subscale

An EFA with an oblique rotation was performed to explore the dimensionality of the Victimization subscale. The first three eigenvalues in descending order were 5.31, 0.80, and 0.65; thus, only the first eigenvalue exceeded 1.0. The differences between successive eigenvalues were 4.52, 0.15, and 0.08, indicating a steep gradient after the first factor. Model fit indices were good for the single-factor solution (RMSEA = 0.035, CFI = 0.981, TLI = 0.975, and SRME = 0.054). Factor loadings were significant for all items (values ranging from 0.61 to 0.81). From these results, we considered the single-factor solution to be optimum for the Victimization subscale, which we determined to be sufficiently unidimensional for IRT analysis.

Witnessing subscale

We performed EFA with an oblique rotation on the Witnessing subscale. The first three eigenvalues in descending order were 5.51, 0.83, and 0.67; thus, only the first eigenvalue exceeded 1.0. The differences between successive eigenvalues were 4.68, 0.15, and 0.21, indicating a steep gradient after the first factor. Model fit indices were good for the single-factor solution (RMSEA = 0.021, CFI = 0.993, TLI = 0.990, and SRMR = 0.052). Factor loadings were significant for all items (values ranging from 0.60 to 0.84). From these results, we considered the single-factor solution also to be optimum for the Witnessing subscale, which was determined to be sufficiently unidimensional for IRT analysis.

Item response theory

Victimization subscale

First, the Victimization subscale of the JaIS with five response categories was analyzed using the graded response model. The estimated slope and category threshold parameters are shown in Table 2. The values of the slope parameters ranged from 1.23 to 2.68, indicating

Table 1. Demographic characteristics of participants (N = 2334)

	Number of students (%)		
	Boys	Girls	Total
Grade			
4th (aged 9–10 years)	219 (9.4%)	198 (8.5%)	417 (17.9%)
5th (aged 10–11 years)	199 (8.5%)	232 (9.9%)	431 (18.5%)
6th (aged 11–12 years)	188 (8.1%)	236 (10.1%)	424 (18.2%)
7th (aged 12–13 years)	175 (7.5%)	178 (7.6%)	353 (15.1%)
8th (aged 13–14 years)	164 (7.0%)	179 (7.7%)	343 (14.7%)
9th (aged 14–15 years)	193 (8.3%)	171 (7.3%)	364 (15.6%)
Total	1138 (48.8%)	1194 (51.2%)	2332 (99.9%)
Language(s) spoken by students (%)			
Japanese	1084 (46.5%)	1133 (48.6%)	2217 (95.1%)
Other language or bilingual	54 (2.3%)	61 (2.6%)	115 (4.9%)

that all items on the Victimization subscale have high discriminating power to classify a respondent as a victim.

The category threshold parameters represent the point along the latent trait (i.e., bullying victimization) at which a respondent has a 0.50 probability of responding above the threshold. That is, a person with a trait (victimization) level of 1.46 has a 50/50 chance of responding *Only once or twice in the past 2 or 3 months*, while a person with a trait level of 1.97 has a 50/50 chance of responding *Two or 3 times a month* in relation to physical bullying.

The item characteristic curve of ‘Physical bullying’ obtained in this procedure is shown in Figure 1a as an example. The respective probabilities of responding *Two or 3 times a month* and *Once a week* were never greater than the other three response options. This result

was true for all nine items and these items thus behaved like a 3-point scale. Therefore, we rearranged the original five categories into the following three categories (0 = *Not being bullied*, 1 = *Only once or less than once a week*, 2 = *Once a week or more*) and reanalyzed the data. The result of this arrangement was found to be the best of all the arrangements of categories because it was the only arrangement in which the probabilities of responding to each category were greater than the others (Fig. 1b).

The item information curves of the Victimization subscale are shown in Figure 1c. The ‘Physical bullying’ item was the most informative at a trait level of approximately 2.0. It was followed by the ‘Lies and false rumors’ item at a trait level of approximately 1.7 and 2.2. For the ‘Cyberbullying’ item, the item information curve was flatter than for other items and had relatively less information.

The test information curves of the Victimization subscale are shown in Figure 1d. We can see that the Victimization subscale provides the most precise measurement for students with a trait level of approximately 2.0 SD, and has little information for students with a trait level of less than 0 SD (i.e., a bullying victimization level of less than the mean score).

Witnessing subscale

The Witnessing subscale with binary response categories was analyzed using the two-parameter logistic model. The item characteristic curve of ‘Witnessing physical bullying’ is shown in Figure 2a as an example, while the estimated slope and location parameters are shown in Table 3. The values of the slope parameters range from 1.37 to 2.73, which indicates that all items in the Witnessing subscale also had a high discriminating power to classify a respondent as an eyewitness. The high discrimination of the items is confirmed by the steep slopes shown in Figure 2a. The location, or difficulty, parameter indicates the trait level required to have a 50% chance of responding *I have seen someone being bullied*. For example, a student with a trait (witnessing bullying) level of 1.72 had a 50/50 chance of responding *I have seen physical bullying*, while a student with a trait level of 1.36 had a 50/50 chance of responding *I have seen verbal bullying*.

The item information curves of the Witnessing subscale are shown in Figure 2b. All items were found to be most informative at a trait level ranging from 1.5 to 2.5 SD. The test information curve of the Witnessing subscale (Fig. 2c) provides the most precise measurement for a person with a trait level of approximately 2.0 SD and little

Table 2. Category thresholds and slope estimates, means and SD for the Victimization subscale of the Japan Ijime Scale: Graded response model

Item	Slope parameter (SE)	Category threshold parameter				Mean	SD
		Between 0 and 1 (SE)	Between 1 and 2 (SE)	Between 2 and 3 (SE)	Between 3 and 4 (SE)		
Physical bullying	2.31	1.46	1.97	2.32	2.7	0.22	0.71
Verbal bullying	2.68	1.02	1.55	1.79	2.07	0.4	0.97
Social exclusion or isolation	1.91	1.76	2.55	2.84	3.3	0.15	0.56
Having money or things taken or damaged	1.89	1.99	3.04	3.4	3.73	0.09	0.41
Lies and false rumors	2.59	1.36	2.15	2.44	2.78	0.19	0.61
Being forced to do things	2.22	1.92	2.6	2.96	3.54	0.1	0.43
Racial bullying	2.36	1.61	2.2	2.52	2.88	0.16	0.59
Sexual bullying	2.06	1.85	2.45	2.76	3.21	0.13	0.55
Cyberbullying	1.23	3.56	4.74	5.54	5.73	0.03	0.25

Category: 0 = *None*, 1 = *Only once or twice in the past 2 or 3 months*, 2 = *Two or 3 times a month*, 3 = *Once a week*, 4 = *Several times a week*. Slope parameter represents discriminability of the item. A high slope parameter value suggests that an item has high ability to differentiate students who have been bullied. Category threshold parameter represents severity of the item (level of victimization), indicating necessity to respond above threshold with a 0.50 level of probability.

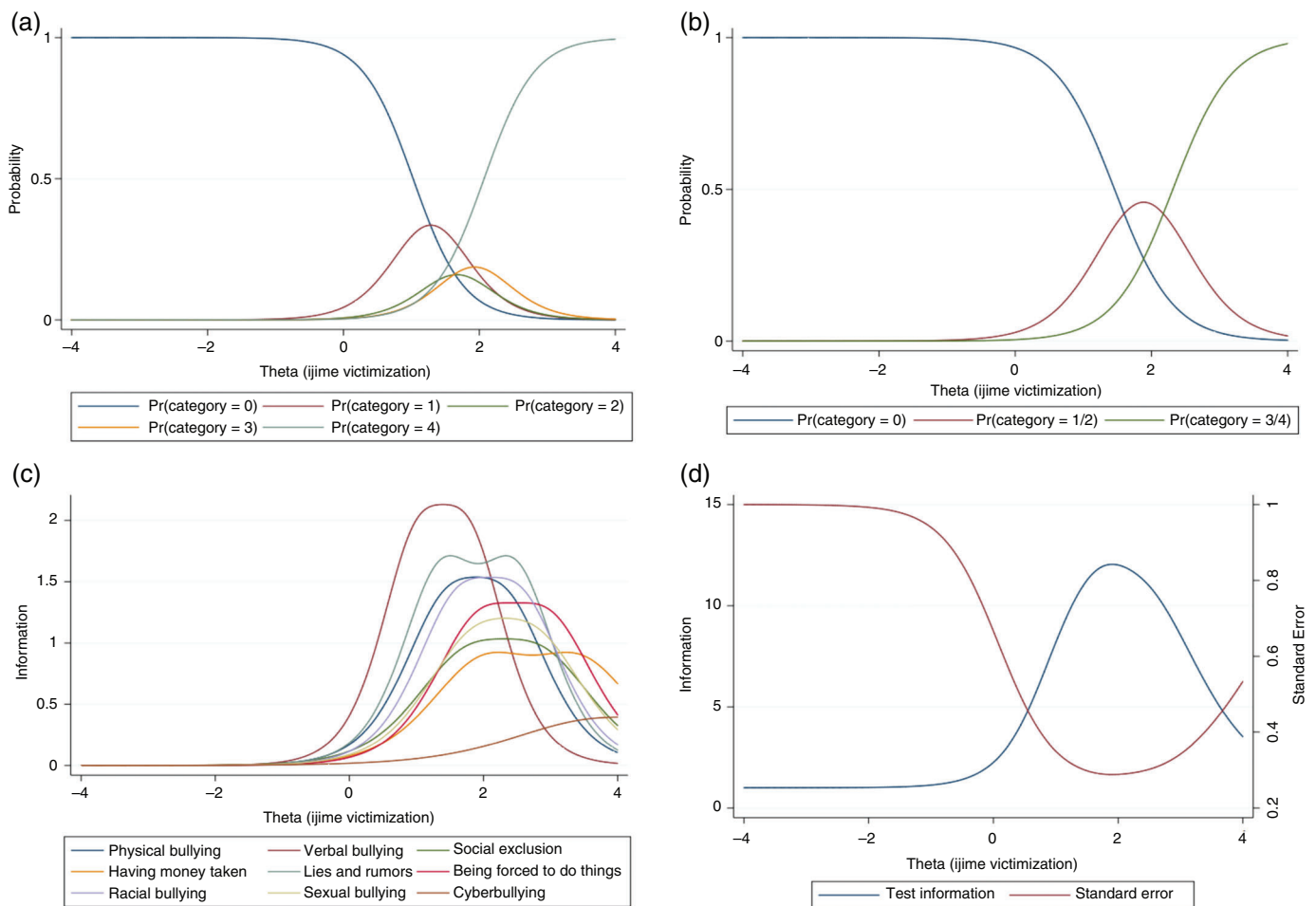


Fig.1 Item response properties for the Victimization subscale of the Japan Ijime Scale. (a) Category characteristic curves of victimization of physical bullying (five response categories). (b) Category characteristic curves of victimization of physical bullying (three response categories). (c) Item information functions of Victimization subscale. (d) Test information functions of Victimization subscale.

information for a person with a trait level of less than 0 SD (i.e., a witnessing level lower than the mean score).

External validity

The association between the Victimization subscale and DSRS-C scores in the adjusted regression analysis was highly significant ($\beta = 0.56$; 95% confidence interval [CI], 0.47–0.65; $P < 0.001$). The association between the Witnessing subscale and DSRS-C scores in the adjusted regression analysis was highly significant ($\beta = 0.32$; 95% CI 0.24–0.40; $P < 0.001$). The association between the Perpetration item and SDQ Conduct Problem scores was also significant ($\beta = 0.63$; 95%CI, 0.51–0.75; $P < 0.001$).

Prevalence of bullying in Japan among fourth to ninth graders

The prevalence of bullying reported in this study is shown in Table 4. The prevalence of bullying victimization, which was reflected in reports of being bullied in some way at least once in the last 2 or 3 months was 35.8% (40.1% in boys vs 31.7% in girls, $\chi^2(1) = 17.9$, $P < 0.001$; 35.6% in students with Japanese nationality vs 39.5% in students with nationalities other than Japanese or dual nationality, $\chi^2(1) = 0.63$, $P = 0.43$; 34.9% in students whose language is Japanese vs 52.2% in students whose language is other than Japanese or bilingual students, $\chi^2(1) = 11.8$, $P = 0.001$). Among nine types of bullying, verbal bullying was reported more frequently, with 19.5% of the total sample having been bullied at least once or twice.

The prevalence of witnessing bullying, as reflected in reports of having seen any type of bullying, was 32.8% (31.9% in boys; 33.7% in girls, $\chi^2(1) = 0.86$, $P = 0.35$), and it was most frequent for verbal bullying (19.5%) among all types of bullying.

The prevalence of bullying perpetration, which was reflected in students' reports of having bullied someone at least once in the last 2 or 3 months, was 11.8% (15.9% in boys; 8.0% in girls), including bully only (3.5%) and bully/victim (8.3%).

Discussion

The JaIS measures bullying victimization, witnessing, and perpetration among school-aged children in Japan. After confirming the unidimensionality of the Victimization and Witnessing subscales, we confirmed that each item of the JaIS Victimization and Witnessing subscales has validity and reliability in terms of satisfactory discriminability, informativity, and precision. In addition, the total scores for the Victimization and Witnessing subscales were highly associated with DSRS-C scores for depression. This indicates that both subscales are in line with external properties. Similarly, the total score for the Perpetration item was highly associated with the SDQ Conduct Problem score; thus, the external validity of the JaIS was well confirmed. In our sample, 35.8% of students reported being bullied in some way at least once in the last 2 or 3 months, while 16.6% of students had been bullied at least two or three times a month. It is important to note that the prevalence was highest in verbal bullying victimization than in the other nine types of bullying victimization, and that the

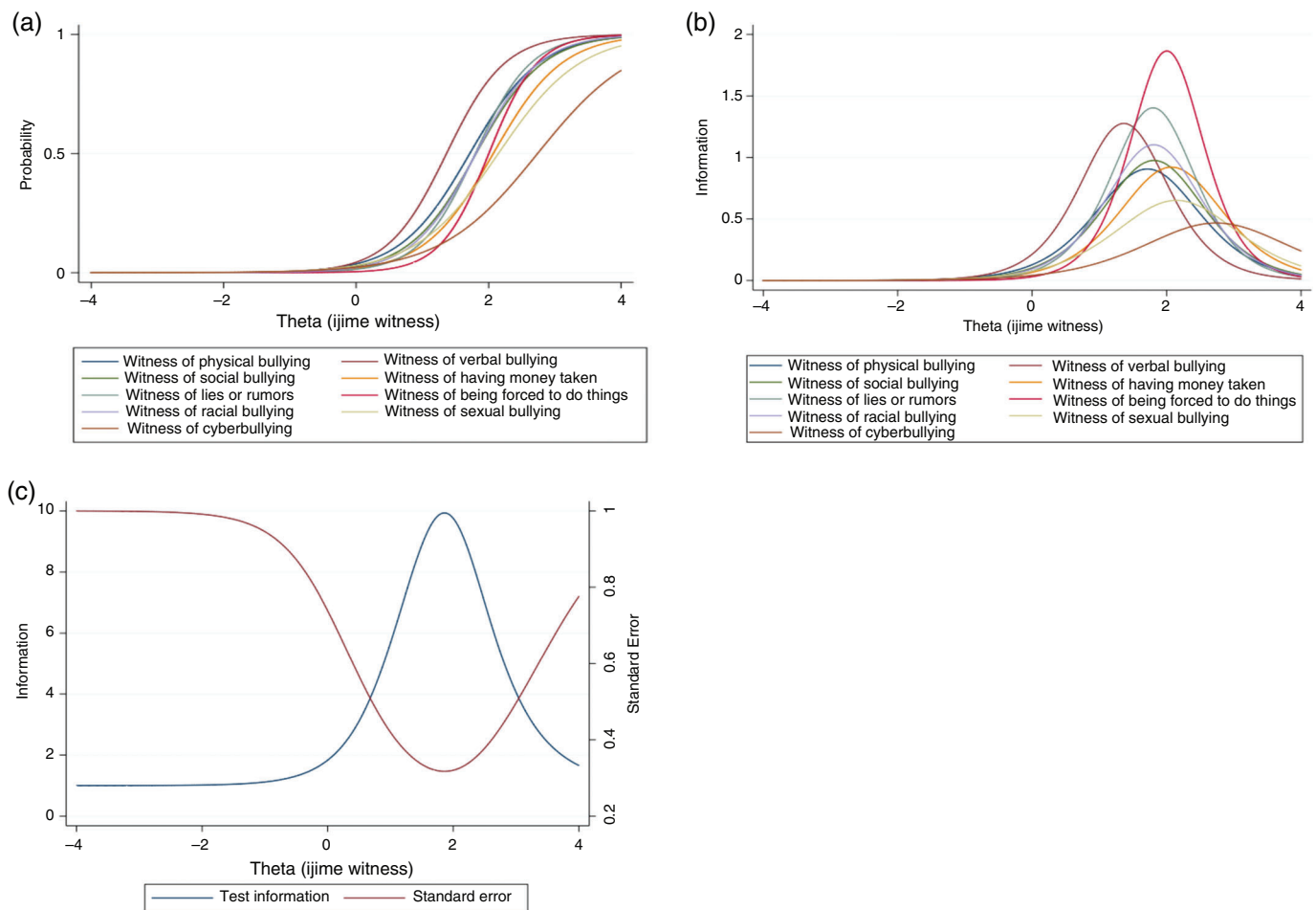


Fig.2 Item response properties for the Witnessing subscale of the Japan Ijime Scale. (a) Item characteristic curves of Witnessing subscale. (b) Item information functions of Witnessing subscale. (c) Test information functions of Witnessing subscale.

prevalence of indirect bullying victimization, such as social exclusion or false rumors, was not frequent.

The results of EFA indicated that single-factor solutions were better suited for both the Victimization and Witnessing subscales. These results suggest that traditional bullying and cyberbullying are explained by the single construct of *bullying* in Japanese elementary schools and junior high schools. However, because there was only one item on cyberbullying, this was unlikely to stand out as one independent factor. In addition, the frequency of cyberbullying was low in this study, so these results should be confirmed using a larger sample, including a higher age group.

Similar to Breivik and Olweus²⁵ and Gonçalves *et al.*,²⁶ our scale's reliability and validity were confirmed using IRT models. In all items of the Victimization subscale, the values of the slope parameters were satisfactorily high and we thus confirmed that the Victimization subscale has sufficient discrimination power (i.e., the scale readily discriminates between students' respective levels of bullying victimization). Breivik and Olweus²⁵ reported that the item 'Being threatened or forced to do things' had the highest discrimination power, while in our study, the 'Verbal bullying' item of the Victimization subscale had the highest discrimination power as shown by the largest slope parameter value in Table 2. Regarding the category threshold parameter, the value was lowest in the 'Verbal bullying' item as with the previous study.²⁵ This result indicates that the threshold level of verbal bullying victimization at which 50% of subjects responded to the item was the lowest among other items.

The Victimization subscale has the highest measurement precision and informativity around +2 SD at theta = 2 from the average

endorsement at theta = 0 (Fig. 1d). This result was consistent with the studies of Breivik and Olweus²⁵ and Gonçalves *et al.*²⁶

Similarly, we were able to confirm that the Witnessing subscale measures the construct of *witnessing bullying* with precision. The factor analysis of the Witnessing subscale showed that it has a single-factor structure like the Victimization subscale. In the IRT analysis of the Witnessing subscale, the item 'Witnessing someone being forced to do things' had the highest discrimination power (the steepest slope as shown in Table 3). The items of 'Witnessing verbal bullying' and 'Witnessing lies and false rumors' also had higher discrimination power compared with other items. The location parameter was lowest in the 'Verbal bullying' item and was shown to be the most informative at around 1–3 SD from the mean. These results are similar to those of the Victimization subscale.

In the examination of external validity, it was confirmed that the higher the victimization score, the higher the depressive mood. The relationship between Witnessing subscale and DSRs-C scores was also confirmed. Furthermore, a higher frequency of perpetration was associated with a higher SDQ Conduct Problem score. These results provide further evidence for the validity of the scale.

However, the frequency of witnessing cyberbullying was relatively low, as was victimization by cyberbullying. Additionally, the informativity (information function by IRT analyses) of cyberbullying was not as high in the Victimization or Witnessing subscales. Thus, we should confirm the factor structure of these subscales using a larger sample, including adolescents.

Table 3. Item location and slope estimates, means and SD for the Witnessing subscale of the Japan Ijime Scale: Two-parameter logistic model

Item	Slope parameter	Location parameter	Mean	SD
Physical bullying	1.90	1.72	0.10	0.30
Verbal bullying	2.26	1.36	0.14	0.35
Social exclusion or isolation	1.98	1.81	0.09	0.28
Having money or things taken or damaged	1.92	2.07	0.06	0.24
Lies and false rumors	2.37	1.80	0.07	0.26
Being forced to do things	2.73	2.00	0.04	0.21
Racial bullying	2.10	1.81	0.08	0.27
Sexual bullying	1.61	2.15	0.07	0.26
Cyberbullying	1.37	2.74	0.05	0.21

Category: 0 = None, 1 = I have seen someone being bullied. Slope parameter represents item discriminability. An item with a higher slope parameter value means that the item has a high ability to differentiate students who have witnessed someone being bullied from those who have not. The location (difficulty) parameter represents difficulty of item, indicating the trait level (i.e., level of witnessing) required to have a 50% chance of endorsing an item. An item with a higher location parameter value means that students with higher latent trait level (i.e., level of witnessing) endorse the item.

According to the examination of the intersections of the item characteristic curves, it was found that the following three response categories were the best of all variations of response categorizations: *Nothing in the past 2 or 3 months*, *Only once or less than once a week in the past 2 or 3 months*, and *About once a week or more in the past 2 or 3 months*. Solberg and Olweus¹⁸ found that the response category *Two or three times a month (in the past couple of months)* is a reasonable and useful lower-bound cut-off point. However, in this study, it was reasonable that the two response categories *Only once or twice in the past 2 or 3 months* and *Two or three times a month in the past 2 or 3 months* were collapsed from inspection of the category characteristic curves of the Victimization subscale. These differences in the wording of the questions among categories may be difficult for younger participants to understand correctly. In future studies, we should examine whether the lower-bound cut-off point differs in Japan compared with Western countries, and whether the wording should be changed.

Investigating the prevalence of bullying victimization using the JaIS, 35.8% of students reported having been bullied in some way at least once in the last 2 or 3 months, while 16.6% of students had been bullied at least two or three times a month. Using the OBVQ, Solberg and Olweus¹⁸ reported a 31.8% bullying prevalence in fourth to ninth graders in Norway for any kind of bullying victimization at least once in the 2 or 3 months prior to the study, but with less frequency than two or three times a month; 10.1% of these students had been bullied at least two or three times a month. In a comparative study, using the OBVQ, Morita¹⁹ reported that the prevalence estimation in Japan was 13.9% in fifth- to ninth-grade Japanese students at least once in the 2 or 3 months before the study, while 8.5% of students had been bullied at least two or three times a month. Although these results may not be directly comparable to our results due to the different methodologies, the differences among the prevalence estimates are surprising. This might be partially due to some historical effects in the awareness of experiences of bullying victimization. Concerning this, a significant decrease in traditional bullying and an increase in cyberbullying has been reported in the literature.⁴⁷ Follow-up surveys using the JaIS in the same region where the current survey was conducted are highly recommended to verify whether traditional bullying has really faded away.

In our study, the prevalence of bullying perpetration in fourth to ninth graders was 11.8%, whereas in Takekawa's study⁴⁸ it was 18.0%. Although the prevalence in this study was lower than that in Takekawa's study,⁴⁸ we should be careful to compare these results since we adopted only one item on bullying perpetration.

In our study, the prevalence of reported witnessing of any type of bullying was 32.8% (31.9% of boys and 33.7% of girls). Rivers *et al.*⁴⁹ reported that 63% of children aged 12 to 16 years reported having witnessed bullying in the UK. In an Australian study, 68.3% of secondary school students had witnessed traditional bullying, cyberbullying, or both in the past 12 months prior to the study.⁵⁰ Although there are few studies on the prevalence of witnessing bullying in Japan, Matsushita³⁸ reported that the proportion of elementary or junior high school students who have witnessed bullying at least once since entering school was 39%. Yonezato³⁴ reported that the prevalence of witnessing bullying among fifth to ninth graders was 40.1% of boys and 48.4% of girls at least once in the last 2 or 3 months prior to the study. The prevalence of witnessing bullying in Japan is thus quite low compared with Western countries. This might be due to Japanese students' lack of education concerning the role of witnesses or bystanders in preventing bullying. In addition, the definition of *ijime*, indicated in the Japanese government's 'Act for the Measures to Prevent Bullying' and adopted nationwide in Japanese schools, emphasizes victims' feelings of physical or mental pain.¹³ Thus, the ambiguity of the government's definition makes it difficult to judge whether one has witnessed bullying or not,¹⁷ because

Table 4. Prevalence of bullying among fourth through ninth graders (N = 2334)

	Physical bullying	Verbal bullying	Social exclusion or isolation	Having money or things taken or damaged	Lies and false rumors	Being forced to do things	Racial bullying	Sexual bullying	Cyberbullying	Total
Frequency of victimization (%)										
<i>Never</i>	2037 (88.0%)	1861 (80.5%)	2093 (90.6%)	2151 (93.1%)	2019 (87.3%)	2163 (93.8%)	2089 (90.6%)	2131 (92.3%)	2257 (97.7%)	1487 (64.2%)
<i>Only once or twice</i>	149 (6.4%)	227 (9.8%)	149 (6.5%)	128 (5.5%)	212 (9.2%)	98 (4.2%)	134 (6.5%)	108 (4.7%)	40 (1.7%)	441 (19.0%)
<i>Two or 3 times a month</i>	57 (2.5%)	72 (3.1%)	25 (1.1%)	14 (0.6%)	36 (1.6%)	21 (0.9%)	36 (1.1%)	29 (1.3%)	8 (0.3%)	127 (5.5%)
<i>Once a week</i>	36 (1.6%)	61 (2.6%)	23 (1.0%)	7 (0.3%)	24 (1.0%)	16 (0.7%)	23 (1.0%)	23 (1.0%)	2 (0.1%)	107 (4.6%)
<i>Several times a week</i>	36 (1.6%)	91 (3.9%)	20 (0.9%)	10 (0.4%)	22 (1.0%)	8 (0.3%)	23 (0.9%)	19 (0.8%)	3 (0.1%)	156 (6.5%)
Total	2315 (100%)	2312 (100%)	2310 (100%)	2310 (100%)	2313 (100%)	2306 (100%)	2305 (100%)	2311 (100%)	2310 (100%)	2318 (100%)
Frequency of witnessing (%)										
<i>Have never seen bullying</i>	2078 (89.9%)	1995 (86.2%)	2115 (91.4%)	2167 (93.7%)	2144 (92.7%)	2208 (95.5%)	2121 (91.9%)	2143 (92.8%)	2201 (95.3%)	1557 (67.2%)
<i>Have seen bullying</i>	234 (10.1%)	320 (13.8%)	200 (8.6%)	145 (6.3%)	169 (7.3%)	104 (4.5%)	188 (8.1%)	167 (7.2%)	109 (4.7%)	760 (32.8%)
Total	2309 (100%)	2315 (100%)	2315 (100%)	2312 (100%)	2313 (100%)	2312 (100%)	2309 (100%)	2310 (100%)	2310 (100%)	2317 (100%)

witnesses cannot judge whether the victim feels physical or mental pain. Some Japanese researchers have argued that the characteristics of *ijime* in Japan differ from those of *bullying* in Western countries.^{51,52} The results of this study that connect higher victimization prevalence and lower witnessing prevalence may reflect a qualitative difference of *ijime* and *bullying* in Western countries, although further evidence is required to draw this conclusion.

One of the limitations of this study is that we did not examine the nine types of bullying perpetration. We included only one item on perpetration and examined the external validity. Therefore, the reliability cannot be examined because the IRT model is bound to be applied to more than one item. Future studies should develop a subscale for bullying perpetration. Another limitation is that, as the measurement of cyberbullying (victimization and witnessing) was based on just one item, cyberbullying was unlikely to be found as an independent factor by the factor analysis. In future studies using a larger sample, cyberbullying victimization may be considered independently.

One of the most important clinical implications of this study is its successful quantification of the need for intervention in instances of school bullying. In Japan, although there have been some studies that have estimated the prevalence of bullying, the definitions and methods have varied and neither the reliability nor the validity of these measures has been thoroughly confirmed. This situation has fostered a social awareness of the need for intervention at an early stage. Repeated measurements of the prevalence of bullying over the same population are also recommended, since it is not well understood whether the structure and frequency of bullying-related experiences would change over time. Further research in this field using the JaIS is awaited.

Acknowledgments

We are grateful to Ms Yoshie Muramatsu, Mr Koichiro Nishikawa, Ms Ryoko Yasuda, Mr Tadashi Nakazato, Ms Taeko Terayama, Ms Kaori Yuge, and Professor Taiichi Katayama of the Institute of Child Developmental Science Research for support in this study. This work was supported by JSPS KAKENHI Grant Numbers 16K15558 and 17K01893. We would like to thank Editage (www.editage.jp) for English-language editing.

Disclosure statement

The authors declare no conflicts of interest.

Author contributions

Conception and design of the study: Y.O., T.N., M.W., and K.J.T. Acquisition and analysis of data: Y.O., T.N., M.W., and N.T. Drafting the manuscript or figures: Y.O., T.N., and K.J.T.

References

1. Arslan S, Hallett V, Akkas E, Akkas OA. Bullying and victimization among Turkish children and adolescents: Examining prevalence and associated health symptoms. *Eur. J. Pediatr.* 2012; **171**: 1549–1557.
2. Rigby K. Consequences of bullying in schools. *Can. J. Psychiatry* 2003; **48**: 583–590.
3. Smith PK, Talamelli L, Cowie H, Naylor P, Chauhan P. Profiles of non-victims, escaped victims, continuing victims and new victims of school bullying. *Br. J. Educ. Psychol.* 2004; **74**: 565–581.
4. Rothern C, Head J, Klineberg E, Stansfeld S. Can social support protect bullied adolescents from adverse outcomes? A prospective study on the effects of bullying on the educational achievement and mental health of adolescents at secondary schools in East London. *J. Adolesc.* 2011; **34**: 579–588.
5. Glew GM, Fan MY, Katon W, Rivara FP, Kernic MA. Bullying, psychosocial adjustment, and academic performance in elementary school. *Arch. Pediatr. Adolesc. Med.* 2005; **159**: 1026–1031.
6. Olweus D. School bullying: Development and some important challenges. *Annu. Rev. Clin. Psychol.* 2013; **9**: 751–780.
7. van Dam DS, van der Ven E, Velthorst E, Selten JP, Morgan C, de Haan L. Childhood bullying and the association with psychosis in non-clinical and clinical samples: A review and meta-analysis. *Psychol. Med.* 2012; **42**: 2463–2474.

8. Copeland WE, Wolke D, Angold A, Costello EJ. Adult psychiatric outcomes of bullying and being bullied by peers in childhood and adolescence. *JAMA Psychiatry* 2013; **70**: 419–426.
9. Lereya ST, Copeland WE, Costello EJ, Wolke D. Adult mental health consequences of peer bullying and maltreatment in childhood: Two cohorts in two countries. *Lancet Psychiatry* 2015; **2**: 524–531.
10. Bowes L, Joinson C, Wolke D, Lewis G. Peer victimisation during adolescence and its impact on depression in early adulthood: Prospective cohort study in the United Kingdom. *BMJ* 2015; **350**: h2469.
11. Klomek AB, Sourander A, Niemelä S *et al.* Childhood bullying behaviors as a risk for suicide attempts and completed suicides: A population-based birth cohort study. *J. Am. Acad. Child Adolesc. Psychiatry* 2009; **48**: 254–261.
12. Shireen F, Janapana H, Rehmatullah S, Temuri H, Azim F. Trauma experience of youngsters and teens: A key issue in suicidal behavior among victims of bullying? *Pak. J. Med. Sci.* 2014; **30**: 206–210.
13. Ministry of Education, Culture, Sports, Science and Technology. Bullying Prevention Measures Promotion Act. [Cited 23 April 2018.] Available from URL: http://www.mext.go.jp/b_menu/shingi/chukyo/chukyo1/gijiroku/attach/1337765.htm
14. Morita Y, Kiyonaga K. *Ijime Kyoushitunoyamai*. Kanekoshobou, Tokyo, 1986 (in Japanese).
15. Taki M. "Ijime"kouino hassei/suijokuyounikansuru jissyoutekikenkyu. *Educ. Stud. Jpn.* 1992; **59**: 1 (in Japanese).
16. National Institute for Educational Policy Research. Ministry of Education, Culture, Sports, Science and Technology in Japan. Ijime Follow-up Study 2013–2015. 2016. [Cited 23 April 2018.] Available from URL: https://www.nier.go.jp/shido/centerhp/2806sien/tuisseki2013-2015_3.pdf
17. Kobayashi E, Miwa S. Comments on studies about bullying: About the definition and measures. *Ibaraki-Daigaku Kyouiku Jissen Kenkyu.* 2013; **32**: 163–174 (in Japanese).
18. Solberg ME, Olweus D. Prevalence estimation of school bullying with the Olweus Bully/Victim Questionnaire. *Aggress. Behav.* 2003; **29**: 239–268.
19. Morita Y. Bullying victimization. In: Morita Y (ed.). *Cross-National Comparative Study of Bullying: A Report from Japan, England, the Netherlands, and Norway*. Kanekoshobou, Tokyo, 2001; 31–54 (in Japanese).
20. Murayama Y, Ito H, Hamada M *et al.* Relationships of bullying behaviors and peer victimization with internalizing/externalizing problems. *Jpn. J. Dev. Psychol.* 2015; **26**: 13–22 (in Japanese).
21. Tochigi M, Nishida A, Shimodera S *et al.* Irregular bedtime and nocturnal cellular phone usage as risk factors for being involved in bullying. *PLoS One* 2012; **7**: e45736.
22. Mizuta A, Okada E, Nakamura M, Yanaguchi H, Ojima T. Association between the time perspective and type of involvement in bullying among adolescents: A cross-sectional study in Japan. *Jpn. J. Nurs. Sci.* 2018; **15**: 156–166.
23. Shimoda Y. *A review of research on bullying among elementary and junior high school students in Japan*. Mem. Fac. Hum. Dev. Univ. Toyama 2014; **8**: 3–37 (in Japanese).
24. Olweus D. *The Revised Olweus Bully/Victim Questionnaire*. Research Center for Health Promotion, University of Bergen. Bergen, Norway, 1996.
25. Breivik K, Olweus D. An item response theory analysis of the Olweus Bullying Scale. *Aggress. Behav.* 2015; **41**: 1–13.
26. Gonçalves FG, Heldt E, Peixoto BN, Rodrigues GA, Filipetto M, Guimarães LSP. Construct validity and reliability of Olweus Bully/Victim Questionnaire – Brazilian version. *Psicol. Reflex. Crit.* 2016; **29**: 27.
27. Olweus D. Invited expert discussion paper. Cyberbullying: An overrated phenomenon? *Eur. J. Dev. Psychol.* 2012; **9**: 520–538.
28. Kubiszewski V, Fontaine R, Potard C, Auzoult L. Does cyberbullying overlap with school bullying when taking modality of involvement into account? *Comput. Hum. Behav.* 2015; **43**: 49–57.
29. Ogi N. *Ijime boushi jissen puroguramu*. Gakuyo Shobo, Tokyo, 1997 in Japanese.
30. Hishida K, Kawabata T, Song S *et al.* Relationships among the consequences of bullying, resiliency, social support and life skills (II) based on a questionnaire survey at eight junior high schools in Niigata and Hiroshima. *Jpn. J. School Health* 2012; **53**: 509–526 (in Japanese).
31. Ishikawa Y. The actual conditions of damages by bullying: Some analyses of data from investigation into the actual conditions of bullying and the opinions about bullying among public junior high school students in Osaka. *Hum. Sci. Res. Bull.* 2010; **9**: 155–184 (in Japanese).
32. Bonds M. *Bully-Proofing Your School: A Comprehensive Approach for Middle Schools*. Sopris West, Longmont, CO, 2000.
33. Salmivalli C, Lappalainen M, Lagerspetz K. Stability and change of involvement in connection with bullying in schools: A two-year follow-up. *Aggress. Behav.* 1998; **24**: 205–218.

34. Yonezato S. Bullying witness. In: Morita Y (ed.). *Cross-National Comparative Study of Bullying: A Report from Japan, England, the Netherlands, and Norway*. Kanekoshobou, Tokyo, 2001; 145–158 (in Japanese).
35. Twemlow SW, Fonagy P, Sacco FC. The role of the bystander in the social architecture of bullying and violence in schools and communities. *Ann. N. Y. Acad. Sci.* 2004; **1036**: 215–232.
36. Polanin JR, Espelage DL, Pigott TD. A meta-analysis of school-based bullying prevention programs' effects on bystander intervention behavior. *Sch. Psychol. Rev.* 2012; **41**: 47–65.
37. Toda Y. Bullying research and supporting prevention practice in schools. *Jpn. J. Dev. Psychol.* 2013; **24**: 460–470 (in Japanese).
38. Matsushita K. Issues of moral education and human rights education in the results of the Survey on Bullying of Children. *J. Fac. Educ. Saga Univ.* 2015; **19**: 37–51 (in Japanese).
39. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct. Equ. Model.* 1999; **6**: 1–55.
40. Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. *Med. Care* 2000; **38**(Suppl. 9): II28–II42.
41. Yang FM, Kao ST. Item response theory for measurement validity. *Shanghai Arch. Psychiatry* 2014; **26**: 171–177.
42. Birnbaum A. *Some latent trait models and their use in inferring an examinee's ability*. In: Lord FM, Novick MR (eds). *Statistical Theories of Mental Test Scores*. Addison–Wesley, Boston, MA, 1968; 397–472.
43. Samejima F. The graded response model. In: van der Linden WJ, Hambleton R (eds). *Handbook of Modern Item Response Theory*. Springer Science+Business Media, New York, NY, 1996; 85–100.
44. Fraley RC, Waller NG, Brennan KA. An item response theory analysis of self-report measures of adult attachment. *J. Pers. Soc. Psychol.* 2000; **78**: 350–365.
45. Namikawa T, Tani I, Wakita T *et al.* Development of a short-form Birleson Depression Self-Rating Scale for Children. *Clinic. Psychiatry* 2011; **53**: 489–496 (in Japanese).
46. Goodman R, Meltzer H, Bailey V. The Strengths and Difficulties Questionnaire: A pilot study on the validity of the self-report version. *Eur. Child. Adolesc. Psychiatry* 1998; **7**: 125–130.
47. Rigby K, Smith PK. Is school bullying really on the rise? *Soc. Psychol. Educ.* 2011; **14**: 441–455.
48. Takekawa I. Bullying perpetration. In: Morita Y (ed.). *Cross-National Comparative Study of Bullying: A Report from Japan, England, the Netherlands, and Norway*. Kanekoshobou, Tokyo, 2001; 159–173 (in Japanese).
49. Rivers I, Poteat VP, Noret N, Ashurst N. Observing bullying at school: The mental health implications of witness status. *Sch. Psychol. Q.* 2009; **24**: 211–223.
50. Roslyn Q, Marilyn C. On standby? A comparison of online and offline witnesses to bullying and their bystander behaviour. *Educ. Psychol.* 2015; **35**: 430–448.
51. Taki M. *Japanese School Bullying: Ijime*. Queen's University, Toronto, Canada, 2001. [Cited 23 April 2018.] Available from URL: <https://www.nier.go.jp/a000110/Toronto.pdf>.
52. Kanetsuna T. Insights into Japanese ijime through a comparison with bullying in England: Focusing on children's group formation and their negative emotions against victims. *Emot. Stud.* 2015; **1**: 17–22 (in Japanese).