Introduction
Empathy facilitates understanding of others’ minds and motivates altruistic behavior (e.g., Decety & Svetlova, 2012). The Interpersonal Reactivity Index (IRI; Davis, 1980) is one of the most widely used measures of empathic traits. This scale measures empathic traits across multiple dimensions: affective scales (Personal Distress [PD] and Empathic Concern [EC]) and cognitive scales (Perspective Taking [PT] and Fantasy [FS]). Although researchers developed a Japanese version of the IRI (Sakurai, 1988; Aketa, 1999), the validity of these translated scales suffered from some limitations. For instance, previous research lacked a test of the content validity of the translated scales. Specifically, they did not confirm that the meaning of each item in Japanese was equivalent to that of the original version by back-translation. Further, previous research did not fully test the criterion-related validity of their scales. Though the subscales of IRI were classified as affective and cognitive components (Davis, 1980), previous research used only the Questionnaire Measure of Emotional Empathy (QMEE; Mehrabian, & Epstein, 1972) to test the criterion-related validity. Thus, the present study aimed to develop the Japanese version of the IRI anew with tests of its construct validity from various perspectives.

Methods
Respondents: A total of 416 adults who registered for participants pool of a marketing research company (Macromill, Inc.) responded to the present study (208 females; \(M_{age} = 39.61, SD = 11.15\)).

IRI-J: Five graduate students majoring in psychology (from the first to the fifth author) translated all IRI items (Davis, 1980) into Japanese. Then, they decided on the final items through discussion, referring to the items of the version of the IRI previously translated into Japanese (Sakurai, 1988; Aketa, 1999; Kanai, 2013). All items were back-translated by a commercial translation service (Editage, Cactus Communications Pvt. Ltd.), and the sixth author, who is also the author of the original IRI, confirmed whether the meanings of the translated items were equivalent to those of the original items. The final version of this scale (IRI-J) consists of 28 items. Respondents answered each item on a 5-point scale ranging from 1 (“Does not describe me well”) to 5 (“Describes me very well”).

Questionnaires: Six scales were used to test the criterion-related validity of the IRI-J: the QMEE (Mehrabian & Epstein, 1972; Kato & Takagi, 1980), the Literary Response Questionnaire (LRQ; Miall & Kuisken, 1995; Osanai & Okada, 2011), trait anxiety in the State-Trait Anxiety Inventory (STAI; Spielberger, Gorsuch & Lushene, 1970; Shimizu & Imae, 1981), the Rosenberg Self-Esteem Scale (Rosenberg, 1965; Mimura & Griffiths, 2007), the Critical Thinking Ability Scale (CTAS; Kusumi & Hirayama, 2013), and the Buss–Perry Aggression Questionnaire (BPAQ; Buss & Perry, 1992; Ando et al., 1999).

Results and Discussion
Results of confirmatory factor analysis revealed that a four-factor model (i.e., PD, EC, PT, and FS) had a poor fit (CFI = .693, RMSEA = .084). Thus, we tested the correlated traits using the correlated methods model, adding one latent method factor of reverse items (see Wouters et al., 2012). The results revealed that revised model had better fit (CFI = .831, RMSEA = .064). Although the fit of this model was not good enough for the standard criteria, it was close to that of versions of the IRI in other languages (CFI = .81 – .86, RMSEA = .05 – .07; De Corte et al., 2007; Fernández Dufey, & Kramp, 2011; Gillet et al., 2013). Therefore, we assumed that the IRI-J consists of the same four traits as the original scale.

Internal consistency of PD, EC, and FS was sufficient (PD: \(a = .76\); EC: \(a = .77\); FS: \(a = .76\)). Although the internal consistency of PT was low (\(a = .65\)), this value was improved after excluding 2 items (#3 and #15) that were weakly correlated with other items (\(a = .75\)). These items were also excluded by Aketa (1999) due to low item-total correlation. Thus, these items seem inappropriate for measuring PT in Japanese populations.

Next, we conducted correlation analysis to test the criterion-related validity of the IRI-J (Table 1). As these results were consistent with the theoretical predictions, we assumed that the IRI-J can measure empathetic traits, as in the original version.

| Scale | Subscale | 1 | 2 | 3 | 4 |
|-------|----------|---|---|---|---|
| 1 IRI  | PD   | - | - | - | - |
| 2 EC   | .19 | - | - | - | - |
| 3 PT   | -15 | .33 | .34 | .35 | .25 |
| 4 FS   | .26 | .24 | .33 | .19 | .25 |

Note: Values indicate Pearson’s correlation (r) Values of PT: (7 items) / (5 items). *: \(p < .05\), **: \(p < .01\).