International Journal of Instruction *April 2021 ● Vol.14, No.2*
e-ISSN: 1308-1470 ● www.e-iji.net *p-ISSN: 1694-609X*
pp. 919-934

Article submission code: 20200327173250
Received: 27/03/2020 Revision: 04/11/2020
Accepted: 25/11/2020 OnlineFirst: 13/03/2021

Validating an Indonesian Version of the What Is Happening in this Class? (WIHIC) Questionnaire Using a Multidimensional Rasch Model

Wardani Rahayu
State University of Jakarta, Indonesia, *wardani.rahayu@unj.ac.id*

Muhammad Dwirifqi Kharisma Putra
Gadjah Mada University, Indonesia, *muhammad.dwirifqi@mail.ugm.ac.id*

Yuli Rahmawati
State University of Jakarta, Indonesia, *yrahmawati@unj.ac.id*

Bahrul Hayat
Syarif Hidayatullah State Islamic University Jakarta, Indonesia, *bhayat59@yahoo.com*

Rekha Bhan Koul
Curtin University, Australia, *r.koul@curtin.edu.au*

The What Is Happening In this Class? (WIHIC) questionnaire is a multidimensional measure for assessing the classroom learning environment based on student perspectives. The purpose of this study was to validate an Indonesian version of WIHIC and to assess the instrument's psychometric properties by using confirmatory factor analysis (CFA) and the Multidimensional Partial Credit Model (MPCM), one of the family of Polytomous Rasch model. The Indonesian WIHIC was administered to 962 high school students (368 male, 594 female), aged 14-18 years, in Jakarta, Indonesia. Results from the CFA model comparison showed that the 7-dimensional factor structure of WIHIC was satisfactory. We performed MPCM and found that this model fit the data of the Indonesian WIHIC reasonably well. Moreover, the estimated Rasch person separation reliabilities of each subscale showed that WIHIC has good internal consistency (0.87 to 0.96). The Indonesian WIHIC questionnaire appears to be a valid and reliable tool for assessing the classroom learning environment in Indonesian high school students.

Keywords: confirmatory factor analysis, learning environments, multidimensionality, Rasch model, validation, WIHIC questionnaire

INTRODUCTION

2020 is the year of the educational revolution in Indonesia. This is closely linked to the election of a new minister of education and culture in Indonesia. The main policy,

Citation: Rahayu, W., Putra, M. D. K., Rahmawati, Y., Hayat, B., & Koul, R. B. (2021). Validating an Indonesian Version of the What Is Happening in this Class? (WIHIC) Questionnaire Using a Multidimensional Rasch Model. *International Journal of Instruction*, 14(2), 919-934. https://doi.org/10.29333/iji.2021.14252a

Merdeka Belajar (Freedom to Learn), focuses on student performance assessments and classroom teaching plans (Cahya, 2019). The Education and Culture Minister of Indonesia is also terminating the Indonesia National Examination (UN; Ujian Nasional) for primary and high school students in 2021, and the UN will be transformed to the Assessment of Minimum Competency and Survey of Character (AKM; *Asesmen Kompetensi Minimum dan Survei Karakter*), which tests literacy, numeracy, and character (which impacts the classroom learning environment profoundly) (Bhwana, 2019).

To date, the assessment of the classroom learning environment already received government attention since the introduction of the Indonesia National Assessment Program (INAP; Ministry of Education and Culture Republic of Indonesia, 2018). INAP was the preliminary program to develop an AKM to implement in 2021. Some aspects related to the learning environment such as teacher support and student task orientation were measured on INAP, which is the first national assessment combining both cognitive and non-cognitive assessments. Maintaining a positive classroom learning environment was an important competence for Indonesian teacher based on the Academic Qualification and Competency Standards for Indonesian Teachers No. 16 (Ministry of Education and Culture Republic of Indonesia, 2007). In 2021, students will have the opportunity to assess directly the classroom environment in their school by means of the AKM. It is critical to have an instrument for measuring student perception of the classroom learning environment that is in line with present conditions.

Since the late of 1970s, there have been numerous attempts to develop an instrument to assess classroom learning environment (Fraser, 1980). In Indonesia, specifically, studies on learning environment were conducted in the late 1970s and early 1980s in West Sumatra and East Java province (Margianti, 2002), by pioneering work resulting in the new Indonesian instrument based upon the Individualized Classroom Environment Questionnaire and the Classroom Environment Scale (Fraser, Pearce & Azmi, 1982). In 2000, learning environment studies again began to make their appearance in Indonesia (Margianti, 2002). These new studies were affiliated with a research program called Learning Environment Research Group Indonesia (LER-GI), and the studies resulted in adapted instruments such as the Student Perception of Opportunity Competence Development (SPOCD; Rahayu, Putra, Iriyadi, Rahmawati, & Koul, 2020) questionnaire.

In general, one of the most widely used instruments in the field of classroom learning environments is What Is Happening In this Class? (WIHIC; Fraser, McRobbie & Fisher, 1996) questionnaire. WIHIC is a multidimensional measure which covers broad areas in educational assessment and evaluation, where the interpretation of each subscale can provide more information about which aspects can be explored further regarding the measurement of learning environment; WIHIC has received attention from researchers, teachers, school administrators, and school system administrators (Chionh & Fraser, 2009; Dorman, Aldridge & Fraser, 2006; Koul & Fisher, 2006). This also in related to authentic assessment which associates learning with real and complicated situations and contexts, and also based on student practices in which real world performances are

repeated (Olfos & Zulantay, 2007; Svinicki, 2004). To date, based on the literature review, there are no other instruments available in Indonesia which cover such dimensions as WIHIC. WIHIC was previously used in Indonesia before the era of "freedom to learn" (e.g., Margianti, 2002; Wahyudi & Treagust, 2006).

Since its initial development in Australia, WIHIC has been used in many countries (Fraser, McRobbie, & Fisher, 1996), including Singapore (Chionh & Fraser, 2009), Greece (Charalampous & Kokkinos, 2017), India (Koul & Fisher, 2006), Indonesia (Margianti, 2002; Wahyudi & Treagust, 2006), and Korea (Kim, Fisher, & Fraser, 2000). The WIHIC also been used for cross-cultural studies in Australia and Taiwan (Aldridge, Fraser, & Huang, 1999). Its wide use shows WIHIC can be adapted for a variety of different cultures. Since one of the last studies of WIHIC, conducted in 2006 in Indonesia, in relation to its use (Wahyudi & Treagust, 2006), there has been considerable development in the learning environment because of changes in the national examination system, school examinations, curriculum, assessment, and educational policy. Adapting WIHIC to the present conditions provides potential advantages to cover major changes in the educational system in Indonesia.

Based on studies that have found that WIHIC has adequate psychometric characteristics when used in a sample of high school students (Aldridge & Fraser, 2000; Wolf & Fraser, 2008), we adapted WIHIC and administered it to our survey population of high school students. In terms of data analysis methods, the analysis of WIHIC has been tested in previous studies using applied qualitative methods (Charalampous & Kokkinos, 2017; Stetson, 2005), as well as quantitative methods such as cluster analysis (Dorman, Aldridge & Fraser, 2006), principal component analysis (Chionh & Fraser, 2009), multitrait-multimethod (Dorman, 2008), exploratory factor analysis (Skordi & Fraser, 2019), and confirmatory factor analysis (Dorman, 2003).

Recommendations from previous research suggest that the methods of statistical analysis in analyzing WIHIC could be more sophisticated than those used in their study; for example, confirmatory factor analysis could be used as well as exploratory factor analysis (Skordi & Fraser, 2019). In general, current psychometric practice encourages the utilize of the Rasch model (Rasch, 1960) to provide a more detailed information about characteristics of items and samples because the Rasch model can provide greater information than just that on the relation between an item and a latent factor (DiStefano, Greer, & Dowdy, 2019; Suryadi, Hayat, & Putra, 2021). The Rasch model has become an important complement to classical test theory (CTT) in the psychometric evaluation of an instrument (Rahayu et al., 2020).

To date, no reports are available investigating the psychometric properties of WIHIC using the Rasch model. The Rasch model is an advanced measurement approach which is able to overcome some limitations of classical test theory (CTT) such as a lack of control over the difficulty level of scale items and appropriate ordering of ordinal response categories (Mitchell-Parker et al., 2018), and the use of Rasch model can overcome limitations from the latest evaluative study of WIHIC (Skordi & Fraser, 2019). To this end, the purpose of this study is to be the first to validate the Indonesian version of WIHIC using a multidimensional Rasch model. In this study, a

multidimensional Rasch model, specifically, the Multidimensional Partial Credit Models (MPCM; Kelderman, 1996) is used to investigate the item fit, item measure, reliability, category functioning, person-item map, and person measure.

METHOD

Participants

Data was gathered from 962 high school students from DKI Jakarta province in Indonesia, consisting of 594 female students and 368 male students with an age range of 14-18 years old (mean age = 16.92; SD = 1.94). A total of 284 students were from grade X, 476 students were from grade XI and 202 students were from grade XII. The students represented 12 public high schools in Jakarta, Indonesia. Data collection was done using an online system. Students were given information about the general aim of the study, and they were assured that their data was handled to protect their privacy. All students participated on a voluntary basis, and no incentive or compensation was provided to them for their participation.

Indonesian Version of WIHIC Questionnaire

The WIHIC questionnaire developed by Fraser, McRobbie, and Fisher (1996) was used to measure activities carried out during classroom learning. The questionnaire consisted of seven subscales containing eight items each, resulting in a total of 56 items. The seven subscales were *Student Cohesiveness* (items 1-8): extent to which students know, help, and support one another; *Teacher Support* (items 9-16): extent to which the teacher helps, befriends, trusts, and shows interest in students; *Involvement* (items 17-24): extent to which students have attentive interest, participate in discussions, do additional work, and enjoy the class; *Investigation* (items 25-32): emphasis on the skills and processes of inquiry and their use in problem solving and investigation; *Task Orientation* (items 33-40): extent to which it is important to complete activities planned and to stay on the subject matter; *Cooperation* (items 41-48): extent to which students cooperate rather than compete with one another on learning tasks; and *Equity* (items 49-56): extent to which students are treated equally by the teacher (Chionh & Fraser, 2009). Each item of the WIHIC questionnaire statement is positively worded and uses a five-point Likert scale: *almost never, seldom, sometimes, often*, and *almost always*.

In adapting the WIHIC Questionnaire to an Indonesian language, we referred to the procedures described in the "*Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures"* (Beaton et al., 2000). We conducted five stages of the adaptation process, which were: (1) *Initial translation*: we translated the original scale into Indonesian. (2) *Synthesis of translations*: we considered cultural factors in choosing the translation results. (3) *Back translation*: the translation was translated back into the initial language scale to see whether there were differences in meaning or not when the scale in Indonesian was translated back into the initial language. (4) *Expert committee*: after correcting the translation by considering the results of the back translation, we discussed the results with content experts. (5) *Test of the prefinal version*: the prefinal scale, through the results of the discussion, was tested with several respondents. This aimed to check if the scale that had been adapted could be understood.

Data Analysis Procedure

Confirmatory Factor Analysis (CFA)

The first analysis was examining factor structure of the Indonesian WIHIC using confirmatory factor analysis. CFA has long been a useful tool for exploring the theoretical dimensions of measurement instrument in educational and psychological research (Harrell-Williams & Wolfe, 2013; Tungkunanan, 2020). Despite consistent findings of WIHIC's multidimensional factor structure, there is no single study that found a unidimensional factor structure of WIHIC. We performed CFA to confirm if the original factor structure holds in our sample. We used the Bayesian estimation of the CFA model (BAYES estimator) since this estimation method has practical advantages (Putra, Rahayu, & Umar, 2019), and in the last decade or so, Bayesian methods have become vastly more popular in nearly all scientific fields (van de Schoot et al., 2017).

In using the Bayesian CFA, we used several statistics and fit indices, including the posterior predictive p-value (PPP-value), the Bayesian Root Mean Square Error of Approximation (BRMSEA), the Bayesian Comparative Fit Index (BCFI), and the Bayesian Tucker–Lewis Index (BTLI) (Garnier-Villareal & Jorgensen, 2020; Hoofs et al., 2018). The proposed Bayesian fit indices allow overall model-fit evaluation using familiar metrics of original indices (Garnier-Villareal & Jorgensen, 2020). The criteria for good fit were: CFI > 0.90, TLI > 0.90, and RMSEA < 0.05 (Wang & Wang, 2019). We proposed two hypothesized models to compare: a one-factor model and a seven-factor model. We hypothesized that the seven-factor model will result in a better fit compare to the one-factor model. The CFA model in this study was analyzed using the Mplus 8.4 program.

Multidimensional Rasch Model

The Rasch measurement model, developed from a simple logistic model (Rasch, 1960) by a group of experts at the University of Chicago, has become a very useful method of constructing a measurement (e.g., Wright, 1977; Wright & Masters, 1982; Wright & Stone, 1979). In general, the Rasch measurement model refers to a family of mathematical models that compute the probability of a certain response to each item given the amount of the latent construct the individual possesses (trait level) and the relation between an item and the construct (item difficulty). The Rasch model scales both persons and items according to the strength of an individual's relation with the latent construct. This model produces measures for each person and each item on a common, interval level scale, called a logit scale (DiStefano, Greer, & Dowdy, 2019; Wright & Stone, 1979).

However, a key assumption in applying the Rasch model is unidimensionality (Reckase, 2009). Since the development of WIHIC, this scale has a multidimensional factor structure with many items with a polytomous scored response. Thus, the unidimensional version of the polytomous Rasch model (e.g., the rating scale model; RSM; Andrich, 1978) or partial credit model (Masters, 1982) was not appropriate because of the multidimensional nature of WIHIC. There was a generalization of the Rasch model called a multidimensional random coefficient of multinomial logit model (MRCMLM;

Adams, Wilson, & Wang, 1997), also called the Multidimensional Rasch Model (e.g., Shih et al., 2013), which can accommodate the multidimensional factor structure. In this study, we used the Multidimensional Partial Credit Model (MPCM; Kelderman, 1996), a form of the MRCMLM, to estimate item and person parameters. The formula for MPCM is as follows:

$$P(u_{ij} = \mathrm{k}|\theta_{\mathrm{j}}) = \frac{e^{\sum_{\ell=1}^{k}(\theta_{j\ell} - b_{i\ell k})}}{1 + \sum_{r=1}^{K_i} e^{(\theta_{j\ell} - b_{i\ell r})}}$$

where $b_{i\ell k}$ is the difficulty parameter for item i on dimension ℓ for score category k, and θ_{j} is the trait level of person (Reckase, 2009). There is an indeterminacy in the estimation of the $b_{i\ell k}$ parameters, so they set parameters equal across the response categories, $k = 1,2, \ldots, K_i$ (Adams, Wilson, & Wang, 1997; Reckase, 2009). The statistical test criteria describing the fit items will use Infit and Outfit mean square statistics when the value is in the range of 0.5-1.5 indicating acceptable fit criteria. The values outside of these bounds may suggest a lack of fit between the item and model (DiStefano, Greer, & Dowdy, 2019). In this study, MPCM analysis was performed using ACER Conquest 4.13 using the Monte Carlo-based estimation method.

FINDINGS

Factor Structure: Dimensionality

The result of the Bayesian CFA indicated that the seven-factor model, according to the original structure of the scale (see, Skordi & Fraser, 2019), is confirmed because the values of the indices are above the acceptable threshold (PPP-value = 0.000; BRMSEA = 0.049, BCFI = 0.904, BTLI = 0.899), compared to the one-factor model (PPP-value = 0.000; BRMSEA = 0.091, BCFI = 0.663, BTLI = 0.652). Thus, based on the model comparison, BRMSEA, and BCFI, the results indicated that the seven-factor model was satisfactory and provided representations of the underlying structure of Indonesian WIHIC. All items loaded significantly (ranging from 0.526 to 0.853) in relation to each latent factor.

Item Measure, Fit Statistics, and Step Parameter

Based on the multidimensional factor structure evidence from the confirmatory factor analysis (CFA), the calibration of the Indonesian Language version of WIHIC items was done using the Multidimensional Partial Credit Model (MPCM). Table 1 shows the results of the calibration of the Indonesian WIHIC, including item measure, Rasch fit statistics, and step (threshold) parameter.

Table 1
Item parameter and fit statistics of Indonesian WIHIC

Item	Dimension	Measure	Infit	Outfit	Step 1	Step 2	Step 3	Step 4
1	Student Cohesiveness	-0.116	1.10	1.10	-2.103	-0.747	0.780	2.070
2		-0.382	1.02	1.03	-0.996	0.080	0.103	0.813
3		-0.413	1.10	1.10	-1.498	-1.481	0.664	2.313
4		-0.534	1.04	1.05	-1.701	-0.573	0.725	1.549
5		0.487	0.88	0.91	-1.326	-1.128	0.475	1.979
6		0.456	1.03	1.05	-1.929	-1.587	1.119	2.397
7		0.639	0.92	0.92	-2.014	-1.295	0.857	2.452
8		-0.036	1.01	1.05	-2.813	-0.746	0.755	2.804
9	Teacher Support	-0.487	0.81	0.83	-2.311	-0.969	0.872	2.408
10		-0.891	0.79	0.82	-2.445	-0.589	0.702	2.333
11		0.407	0.88	0.88	-1.622	-0.828	0.785	1.665
12		-0.566	0.77	0.81	-2.030	-0.662	0.796	1.896
13		-0.619	1.09	1.06	-3.077	-0.527	0.983	2.621
14		0.483	1.02	0.98	-1.689	-0.777	0.621	1.845
15		1.929	1.48	1.35	-1.201	-0.610	0.719	1.092
16		-0.255	1.15	1.08	-1.778	-1.329	0.843	2.264
17	Investigation	-0.234	0.89	0.90	-2.339	-0.709	0.702	2.346
18		-0.288	0.87	0.87	-2.198	-0.910	0.735	2.373
19		0.265	1.33	1.31	-2.686	-1.477	1.176	2.986
20		0.392	0.86	0.87	-2.360	-1.313	1.041	2.633
21		-0.185	1.10	1.09	-3.113	-0.936	0.986	3.063
22		-0.168	0.85	0.84	-2.365	-1.031	0.693	2.704
23		-0.278	0.83	0.85	-2.134	-1.255	0.548	2.842
24		0.496	1.00	0.99	-2.102	-1.215	0.846	2.471
25	Involvement	0.196	0.87	0.88	-2.128	-0.875	0.704	2.299
26		0.173	1.00	0.99	-2.077	-0.960	0.555	2.482
27		0.262	0.85	0.85	-2.130	-1.058	0.669	2.519
28		0.863	1.07	1.05	-2.260	-1.089	0.793	2.557
29		-0.220	1.02	0.99	-1.987	-0.809	0.531	2.264
30		-0.333	0.80	0.82	-2.514	-0.897	0.865	2.546
31		-0.340	1.05	1.01	-2.362	-1.241	0.762	2.841
32		-0.601	1.04	1.07	-2.403	-1.045	0.555	2.893
33	Task Orientation	-0.150	1.12	1.08	-1.149	-1.026	0.448	1.828
34		0.373	0.91	0.83	-1.344	-0.754	0.301	1.797
35		0.253	0.91	0.87	-0.933	-0.853	0.191	1.595
36		-0.170	1.11	1.14	-0.952	-0.905	0.360	1.498
37		-0.046	0.92	0.98	-1.023	-0.598	0.230	1.391
38		0.493	1.06	1.03	-1.762	-1.142	0.586	2.318
39		-0.639	0.83	0.91	-1.463	-0.937	0.418	1.982
40		-0.115	0.90	0.93	-1.409	-1.330	0.640	1.998
41	Cooperation	-0.483	1.34	1.23	-2.711	-1.278	0.976	3.013
42		0.208	1.23	1.20	-2.174	-0.963	0.604	2.533
43		-0.757	0.99	1.03	-2.009	-1.193	0.760	2.442
44		-0.124	0.95	0.97	-1.401	-1.396	0.477	2.320
45		0.138	1.25	1.16	-1.614	-1.270	0.454	2.430
46		0.434	1.14	1.14	-1.424	-1.140	0.355	2.209
47		0.078	0.95	0.96	-1.628	-0.986	0.323	2.292
48		0.506	1.00	0.98	-1.416	-1.373	0.686	2.103
49	Equity	0.363	1.42	1.11	-2.406	-0.985	0.779	2.612
50		0.273	0.97	0.97	-2.067	-1.185	0.665	2.587
51		-0.216	1.01	1.01	-2.924	-0.497	0.620	2.802
52		-0.374	0.92	0.99	-2.567	-0.882	0.717	2.731
53		-0.127	0.82	0.91	-2.272	-1.034	0.761	2.545
54		-0.232	0.85	0.91	-2.684	-0.974	0.802	2.856
55		0.772	1.37	1.27	-2.285	-1.388	1.038	2.636
56		-0.458	0.97	1.01	-2.554	-1.369	1.003	2.920

From Table 1, it can be seen that there are no items found as a misfit since all items had the amount of Infit and Outfit in the recommended criteria (0.5-1.5). In sum, all items fit the multidimensional PCM (MPCM). In terms of item measure, the easiest item to endorse was Item 10, which located on the -0.891 logit (*Dalam pelaksanaan pembelajaran, guru saya memberikan bantuan ataupun bimbingan dengan baik*), and by far, the most difficult item to endorse was Item 15, which located on the 1.929 logit (*Guru memberikan motivasi kepada saya baik didalam maupun diluar kelas*). Both of the items measured teacher support aspects. None of the items displayed a disordered threshold (step parameter) since all thresholds were ordered from the lowest to highest value. Based on that evidence, we confirmed that the Indonesian WIHIC items and response categories functioned well.

Correlation between Dimensions: MPCM and CFA Results

As shown in the lower triangle of Table 2, the direct estimates of the correlation matrix for the seven subscales of WIHIC using the multidimensional Rasch analysis approach were between 0.667 and 0.994. In general, the pattern of the correlations was consistent with the theory of classroom learning environment, where all of the positive dimensions about classroom learning environment are highly correlated to each other, supporting the expectation of the outcome of WIHIC. All correlations were statistically significant.

Table 2
Latent correlation between WIHIC's subscale: MPCM and IFA

Dimensions	SC	TS	INVL	INVS	TOR	COP	EQU
SC	1	0.717	0.759	0.649	0.700	0.747	0.694
TS	0.840	1	0.750	0.676	0.629	0.606	0.746
INVL	0.806	0.838	1	0.861	0.677	0.665	0.621
INVS	0.756	0.740	0.944	1	0.599	0.684	0.643
TOR	0.853	0.708	0.736	0.815	1	0.635	0.703
COP	0.866	0.667	0.750	0.777	0.787	1	0.677
EQU	0.797	0.836	0.696	0.687	0.805	0.755	1

Note: SC: students cohesiveness, TS: teacher-support, INVL: involvements, INVS: investigation, TOR: task orientation, COP: cooperation, EQU: equity

As shown in the upper triangle of Table 2, the direct estimates of the correlation of the correlation matrix for the seven subscales of WIHIC using the CFA approach were between 0.565 and 0.867. In general, the pattern of the positive direction of the correlations was consistent with the theory of classroom learning environment and in parallel also consistent with findings from multidimensional Rasch analysis. Using these two methods, we found the Indonesian WIHIC are in line with original WIHIC.

Wright Maps and Reliability

After exposure to information about the results of item parameter estimation, the link between the levels of the "latent trait" of the person and also the level of "difficulty" of the items can be compared directly using the Wright Map (Wilson & Draney, 2002). In MPCM, item parameter and person parameters were calibrated to be on the same metric so that within a single dimension all model parameter estimates can be compared on the

same scales (Liu, Wilson, & Paek, 2008). Because of the multidimensional structure of WIHIC, the resulting Wright Map has a description of the level of ability of each of the seven dimensions, resulting in seven ability distributions and one item parameter distribution (See Figure 1).

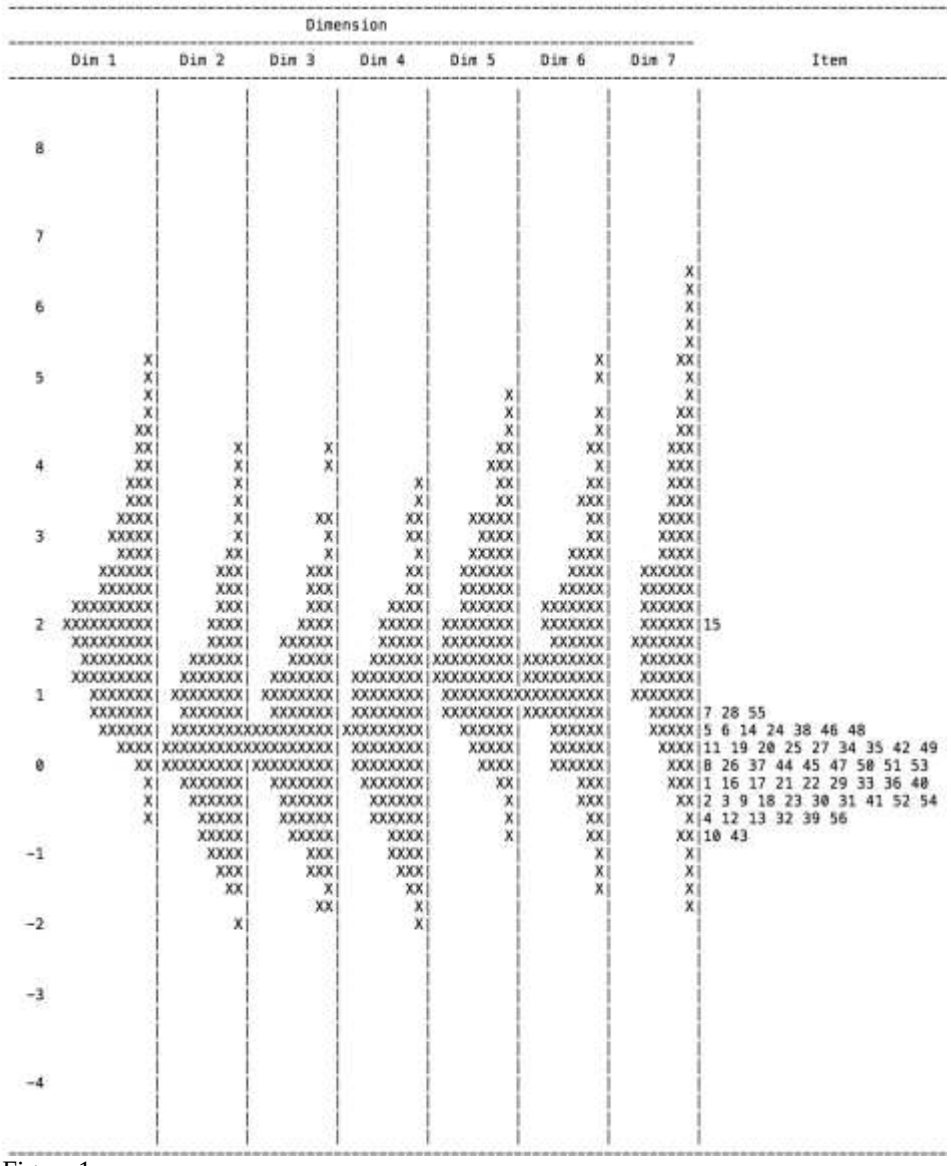

Figure 1
Wright Map of Indonesian WIHIC

In Figure 1, it can be seen that the mean of Rasch person measures in each dimension of WIHIC varies. This shows that the distinctiveness of the dimensionality structure of WIHIC is multidimensional. In addition, the findings show that in all dimensions, the distribution had the highest minimum value in the value of -2 logit, where the distribution tended to lead to a high positive perception of classroom learning environment in a logit scale. All items seem easy to most of the respondents, which tend to endorse high response category.

Table 3
Mean of person measure and reliability

Dimension	Person Mean	Range		EAP/PV reliability
		Minimum	Maximum	
Student Cohesiveness	2.08	-0.71	5.32	0.95
Teacher Support	0.32	-2.09	4.28	0.92
Investigation	0.38	-1.91	4.28	0.96
Involvement	0.93	-2.09	3.87	0.87
Task Orientation	1.81	-0.97	4.83	0.93
Cooperation	1.74	-1.72	5.32	0.92
Equity	1.77	-1.91	6.41	0.88

Table 3 provides the average person measure and EAP/PV reliability as a person separation reliability (PSR) based on a Rasch analysis for each WIHIC subscale. According to Table 3, student cohesiveness had the highest mean of a person's ability, indicating that such aspects are the strength of Indonesian student in the context of classroom learning environments. The subscales *student cohesiveness*, *task orientation*, *cooperation*, and *equity* all had a mean above 1.5 logit, indicating these WIHIC subscales were high. Teacher support had the lowest mean, which indicates that, from the student perspective, teacher support was not optimal in the classroom learning environment. Furthermore, the EAP/PV reliability was used as an indicator of the internal consistency of all WIHIC subscales ranging from 0.87 to 0.96, indicating excellent internal consistency, with all coefficients being considerably much higher than the minimum satisfactory value of 0.70 of the Rasch person separation reliabilities (Geldenhuys & Bosch, 2020).

DISCUSSION

The current study reported on Rasch analysis conducted to advance the psychometric validation of an Indonesian version of the WIHIC, a widely used multidimensional measure of classroom learning environments. In this study, we illustrate how multidimensional Rasch analysis can be applied to analyze WIHIC, which has a multidimensional factor structure, when the application of this model is still largely unknown to many practitioners (Cheng, Wang, & Ho, 2009). Findings from our study indicate that the Indonesian WIHIC's psychometric properties were very strong with all items being significant in the CFA model, all items fitting with the MPCM, all response categories functioning well, and all items with high internal consistencies.

Before the Rasch analysis was performed, we used Bayesian CFA to confirm the factor structure of the Indonesian WIHIC. The CFA established the multidimensionality of the

WIHIC with good overall-model-fit indices. This evidence found clear multidimensionality based on Bayesian CFA and supported the use of multidimensional Rasch analysis, confirming the original structure of the scale (see, Aldridge et al., 1999; Skordi & Fraser, 2009). This also followed the suggestion from a previous study of WIHIC that suggested the need for a larger sample size and more sophisticated methods to assess its psychometric properties (Skordi & Fraser, 2009). The previous study also recommended combining the CFA and the Rasch model in order to have more information regarding psychometric characteristics (Rahayu et al., 2020). In this study, we followed these suggestions both by using CFA followed by a Rasch model in validating WIHIC and by having a larger sample size.

In terms of item fit and category functioning, we found that all items fit to the models and all response categories functioned well since there was no disordered threshold. These indicate that the Indonesian WIHIC can function well in the present-day conditions, which is needed since the last study of the Indonesian WIHIC was in the 2000s (e.g., Wahyudi & Treagust, 2006). We also found adequate internal consistency for all subscales (EAP/PV reliability = 0.87 to 0.96), which is consistent with Cronbach's alpha found in the Singapore version of the WIHIC, with an alpha range around 0.90 (Chionh & Fraser, 2009). The latest psychometric evaluation of the original WIHIC revealed an alpha range from 0.90 to 0.95 (Skordi & Fraser, 2019). Based on the EAP/PV reliability, the Indonesian WIHIC has high reliabilities (EAP/PV reliability = 0.87 to 0.96). We conclude that WIHIC has adequate internal consistencies.

From a methodological perspective, with unidimensional assumptions in the simple Rasch model, our research provides an example of the application of Rasch-based analysis tools that appropriate for analyzing multidimensional data. Some studies generally perform a Rasch analysis separately for each dimension (e.g., McCreary et al., 2013; Ogunbode, Henn, & Tausch, 2020), but in this study, we used the Multidimensional Rasch Model and found important information in the form of correlations between latent variables and no correlation between observed scores if performing a separate Rasch analysis for each dimension.

In addition, the multidimensional Rasch analyses were overly supportive of the correlation pattern of the WIHIC, and we found that the seven dimensions of WIHIC correlated to each other with significant positive correlation. In line with previous research (e.g., Chionh & Fraser, 2009; Skordi & Fraser, 2019), our study also confirms the original structure of the WIHIC, with all of the dimensions having a positive environment as theorized (Aldridge et al., 1999). Based on the person trait level mean, we found *student cohesiveness* was the highest aspect that Indonesian students have; conversely, *teacher support* was the lowest compared to the other subscales. Future research should address this problem in more depth and find solutions for intervention and/or for increasing that lowest and maintaining that highest aspect.

Limitations

The following limitations should be noted. First, the Rasch analysis applied in the present research was based on a sample of data collected only in one province and

capital city of Indonesia, which affects the generalization because there are 33 other provinces in Indonesia. Jakarta was also the "best" province in terms of infrastructure, while the students from other provinces, especially in remote areas, tend to experience more difficulties than students in Jakarta. Second, in this research, we used nonprobability sampling, and this technique may not produce an accurate representation of the population in the study areas. We hope the Indonesian WIHIC will be administered in higher populations from Indonesia's vast geographical area.

CONCLUSION

The psychometric properties of the Indonesian WIHIC was examined in the current study. The values of reliability for each subscale of the Indonesian WIHIC were higher when using the MPCM than when using the classical test theory approach (e.g., Cronbach's alpha). However, the findings of this study should be administered in other sample characteristics (e.g., primary school students) to verify that the multidimensional model was able to fit in another type of sample. The Indonesian WIHIC can be used as a recommendation and guideline for the development of the classroom learning environment in the Assessment of Minimum Competency and Survey of Character (AKM) that will be implemented on 2021.

REFERENCES

Adams, R. J., Wilson, M., & Wang, W. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, 21(1), 1-23.

Aldridge, J. M., & Fraser, B. J. (2000). A cross-cultural study of classroom learning environments in Australia and Taiwan. *Learning Environments Research*, 3(2), 101–134.

Aldridge, J. M., Fraser, B. J., & Huang, T.-C.I. (1999). Investigating classroom environments in Taiwan and Australia with multiple research methods. *Journal of Educational Research*, 93(1), 48–62.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43(4), 561–573.

Beaton, D. E., Bombardier C., Guillemin F., & Ferraz M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, 25(24), 3186-91.

Bhwana, P. G. (2019, December 11). Nadiem Makarim announces no more national exam per 2021. *Tempo.com.* https://en.tempo.co/read/1282559/nadiem-makarim-announces-no-more-national-exam-per-2021

Cahya, G. H. (2019, December 15). Teacher quality may hold back Nadiem's reforms. *The Jakarta Post*. https://www.thejakartapost.com/news/2019/12/15/teacher-quality-may-hold-back-nadiems-reforms.html

Charalampous, K., & Kokkinos, C. M. (2017). The "What Is Happening in This Class" questionnaire: a qualitative examination in elementary classrooms. *Journal of Research in Childhood Education*, 31(3), 379–400.

Cheng, Y., Wang, W., & Ho, Y. (2009). Multidimensional rasch analysis of a psychological test with multiple subtest: a statistical solution for the bandwidth-fidelity dilemma. *Educational and Psychological Measurement*, 69(3), 369-388.

Chionh, Y. H., & Fraser, B. J. (2009). Classroom environment, achievement, attitudes and self-esteem in geography and mathematics in Singapore. *International Research in Geographical and Environmental Education*, 18(1), 29-44.

DiStefano, C., Greer, F. W., & Dowdy, E. (2019). Examining the BASC-3 BESS parent form–preschool using Rasch methodology. *Assessment,* 26(6), 1162-1175.

Dorman, J. P. (2003). Cross-national validation of the what is happening in this class? (WIHIC) questionnaire using confirmatory factor analysis. *Learning Environments Research*, 6(3), 231–245.

Dorman, J. P. (2008). Use of multitrait-multimethod modelling to validate actual and preferred forms of the What Is Happening in This Class? (WIHIC) Questionnaire. *Learning Environments Research,* 11(3), 179–193.

Dorman, J. P., Aldridge, J. M., & Fraser, B. J. (2006). Using students' assessment of classroom environment to develop a typology of secondary school classrooms. *International Education Journal,* 7(7), 906-915.

Fraser, B. J. (1980). Research on classroom learning environment in the 1970's and 1980's'. *Studies in Educational Evaluation*, 6(3), 221–223.

Fraser, B. J., McRobbie, C. J., & Fisher, D. L. (1996, April). *Development, validation and use of personal and class forms of a new classroom environment instrument* [Paper presentation]. Annual Meeting of the American Educational Research Association, New York, NY.

Fraser, B. J., Pearse, R., & Azmi. (1982). A study of Indonesian students' perceptions of classroom psychosocial environment. *International Review of Education*, 28(3), 337-355.

Garnier-Villarreal, M., & Jorgensen, T. D. (2020). Adapting fit indices for Bayesian structural equation modeling: comparison to maximum likelihood. *Psychological Methods,* 25(1), 46–70.

Geldenhuys, M. & Bosch, A. (2020). A Rasch adapted version of the 30-item Bem Sex Role Inventory (BSRI). *Journal of Personality Assessment*, 102(3), 428-439.

Harrell-Williams, L. M., & Wolfe, E. W. (2013). The influence of between-dimension correlation, misfit, and test length on multidimensional rasch model information-based fit index accuracy. *Educational and Psychological Measurement,* 73(4), 672-689.

Hoofs, H., van de Schoot, R., Jansen, N. W. H., & Kant, I. (2018). Evaluating model fit in bayesian confirmatory factor analysis with large samples: simulation study introducing the BRMSEA. *Educational and Psychological Measurement*, 78(4), 537-568.

Kelderman, H. (1996). Multidimensional rasch models for partial-credit scoring. *Applied Psychological Measurement*, 20(2), 155-168.

Kim, H. B., Fisher, D. L., & Fraser, B. J. (2000). Classroom environment and teacher interpersonal behaviour in secondary science classes in Korea. *Evaluation and Research in Education*, 14(1), 3–22.

Koul, R., & Fisher, D. (2006). A contemporary study of learning environments in Jammu, India. In D. Fisher & M. S. Khine (Eds.), *Contemporary approaches to research on learning environments worldviews* (pp. 273-296). Singapore: World Scientific Publishing.

Liu, O. L., Wilson, M., & Paek, I. (2008). A multidimensional rasch analysis of gender differences in PISA mathematics. *Journal of Applied Measurement*, 9(1), 18-35.

Margianti, E. S. (2002). Learning environment research in indonesia. In S. C. Goh & M. S. Khine (Eds.), *Studies in educational learning environments* (pp. 153–167). Singapore: World Scientific.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149-174.

McCreary, L. L., Conrad, K. M., Conrad, K. J., Scott, C. K., Funk, R. R., & Dennis, M. L. (2013). Using the rasch measurement model in psychometric analysis of the family effectiveness measure. *Nursing Research*, 62(3), 149-159.

Ministry of Education and Culture Republic of Indonesia. (2007). *Academic qualification and competency standards for Indonesian teachers No. 16.* Jakarta, Indonesia: Ministry of Education and Culture Republic of Indonesia

Ministry of Education and Culture Republic of Indonesia. (2018). *Indonesian National Assessment Programme (INAP).* Available at https://puspendik.kemdikbud.go.id/inap-sd/

Mitchell-Parker, K., Medvedev, O. N., Krageloh, C. U., & Siegert, R. J. (2018). Rasch analysis of the frost multidimensional perfectionism scale. *Australian Journal of Psychology*, 70(3), 258–268.

Ogunbode, C. A., Henn, L., & Tausch, N. (2020) Context-appropriate environmental attitude measurement in Nigeria using the Campbell paradigm. *Environment, Development and Sustainability*, 22(3), 2141–2158.

Olfos, R., & Zulantay, H. (2007). Reliability and validity of authentic assessment in a web based course. *Educational Technological & Society*, 10(4), 156-173.

Putra, M. D. K., Rahayu, W., & Umar, J. (2019). Indonesian-language version of general self-efficacy scale-12 using Bayesian confirmatory factor analysis: a construct validity testing. *Jurnal Penelitian dan Evaluasi Pendidikan*, 23(1), 12-25.

Rahayu, W., Putra, M. D. K., Rahmawati, Y., Iriyadi, D., & Koul, R. B. (2020). A Rasch and factor analysis of an Indonesian version of the Student Perception of

Opportunity Competence Development (SPOCD) Questionnaire. *Cogent Education*, 7(1): 1721633

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen, Denmark: Danish Institute for Educational Research

Reckase, M. D. (2009). *Multidimensional item response theory*. New York, NY: Springer

Shih, C., Chen, C., Sheu, C., Lang, H., & Hsieh, C. (2013). Validating and improving the reliability of the EORTC QLQ-C30 using a multidimensional rasch model. *Value in Health*, 16(5), 848-854.

Skordi, P., & Fraser, B. J. (2019). Validity and use of the What Is Happening In this Class? (WIHIC) questionnaire in university business statistics classrooms. *Learning Environments Research*, 22(2), 275-295.

Suryadi, B., Hayat, B., & Putra, M. D. K. (2021). The Indonesian version of the Life Orientation Test-Revised (LOT-R): psychometric properties based on the rasch model. *Cogent Psychology*, 8(1): 1869375

Tungkunanan, P. (2020). Learning model of undergraduate students: confirmatory factor analysis. *International Journal of Instruction*, 13(3), 665-678.

Svinicki, M. D. (2004). Authentic assessment: testing in reality. *New Directions for Teaching and Learning*, 100, 23-29.

van de Schoot, R., Winter, S. D., Ryan, O., Zondervan-Zwijnenburg, M., & Depaoli, S. (2017). A systematic review of bayesian articles in psychology: the last 25 years. *Psychological Methods*, 22(2), 217-239

Wahyudi., & Treagust, D. F. (2006). Science education in Indonesia: a classroom learning environment perspective. In D. L. Fisher, & M. S. Khine (Eds.), *Contemporary approaches to research on learning environments* (pp. 221–246). Singapore: World Scientific.

Wang, J. & Wang, X. (2019). *Structural equation modeling: applications using mplus* (2nd ed.), Chicester, UK: John Wiley & Sons Ltd

Wilson, M., & Draney, K. (2002). A technique for setting standards and maintaining them over time. In S. Nishisato, Y. Baba, H. Bozdogan, & K. Kanefugi (Eds.), *Measurement and multivariate analysis* (pp. 325-332). Tokyo: Springer-Verlag.

Wolf, S. J., & Fraser, B. J. (2008). Learning environment, attitudes and achievement among middle-school science students using inquiry-based laboratory activities. *Research in Science Education*, 38(3), 321–341.

Wright, B. D. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement*, 14(2), 97-116.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago, IL: MESA Press

Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago, IL: MESA Press