

Validation of the Mental Retardation Attitude Inventory-Revised (MRAI-R): A Multidimensional Rasch Analysis

Ka-Lam Sam, Chunxiao Li, and Sing-Kai Lo

Abstract—The purpose of this study is to examine the psychometric properties of the Mental Retardation Attitude Inventory-Revised (MRAI-R) via a multidimensional random coefficients multinomial logit model (MRCMLM) under the Rasch analysis. A total of 521 college students (294 males; 227 females), aged 18-25, in the area of south-east China, completed the MRAI-R (Chinese version). The results showed that the partial credit model had a better goodness of fit than the rating scale model. Among all 29 items, two of them exhibited gender differences (item 2 and 8), and two disordered in their step analyses (item 2 and 22). With the identified four subconstructs, the subscale Subtle Derogatory Beliefs (SUDB) had a relatively low reliability (0.496) while compared to the other three subscales (0.664-0.833). The current results revealed the statistical feasibility of MRAI-R through item calibration. Further work on investigating the details of the item nature itself is recommended.

Index Terms—Mental retardation, Chinese college students, MRAI-R, multidimensional rasch analysis.

I. INTRODUCTION

Societal inclusion means integrate people with disabilities into the community. In the past few decades, it has been witnessed that there is a global trend for adaptation of inclusive settings [1], [2]. Societal inclusion is of importance to enhance quality of life and to promote psychological wellbeing among people with intellectual disabilities, ID [3]. ID is a modern term of mental retardation, MR, while MR is specified in the medical field.

China is following the global practice of societal inclusion due to the rapid growth of economics [4]. However, there are many challenges regarding the integration of people with ID [4]; and one of those is the negative attitude and discrimination towards people with disabilities, especially for individuals with ID [5].

It is crucial to examine people's attitude towards societal inclusion via a well-validated instrument since the critical role that attitude plays [6]. However, there is lack of instruments for measuring public attitude towards inclusion, and of people with ID in China [7]. Mental Retardation Attitude Inventory-Revised, MRAI-R [8] is one of the commonly used measures for investigating people's attitude towards ID [7]. The MRAI-R contains 29 items, contributing

to four subscales/domains, namely: integration-segregation (INSE), social distance (SDIS), private rights (PRRT), and subtle derogatory belief (SUDB). More specifically, INSE evaluates people's view towards classroom ID inclusion; the SDIS assesses the willingness to live near or be asocial with people with ID; the PRRT measures the belief of the private rights to express views on ID inclusion; and SUDB assesses the perception towards ID individuals [8].

The MRAI-R was devised in the United States and its reliability and validity has been well established in American adults and high school students [8], [9]. Horner-Johnson and his colleagues (2002) examined the psychometric properties of MRAI-R in Japanese college students and it was found that Japanese samples fit for the try-out factor structure of MRAI-R [10]. More recently, Hampton and Xiao (2008) investigated the psychometric properties of MRAI-R in Chinese college students, it was concluded that factor structure of the MRAI-R in Chinese college students could not be replicated exactly what the structure found in American adults [7]. Thus, further examination on the utility of MRAI-R in Chinese context was needed.

The present study aimed to explore the reliability and validity of the MRAI-R in Chinese college students through Rasch analysis. To the best of our knowledge, no previous studies applied Rasch analysis for validation of MRAI-R. Compared with the classic test theory (CTT), item response theory (IRT) viewed the item and person as two independent parameters. While adopting IRT models, the parameters of person's ability and item difficulty are posed on the same scale on item level with a probabilistic distribution of respondents' endorsement. The invariance attributes of person and item stats make IRT analysis superior to the traditional models [11]. For example, the partial credit model usually applies for polytomous item [12], whereas the unidimensional Rasch model (an IRT model) takes the form

$$\log\left(\frac{p_{ijk}}{p_{ij(k-1)}}\right) = \theta_i - b_{jk} \quad (1)$$

where p_{ijk} and $p_{ij(k-1)}$ are the probabilities of scoring category k and $k-1$ to item j for testee i with ability θ_i , respectively, and b_{jk} represents the k^{th} step difficulty of item j .

In general, the unidimensional Rasch analysis presumes that the testees' performances on the full test were determined by a single trait, all the items within the test are expected to measure the same trait. If any item loads on more than one dimension, this item should be considered to modify

Manuscript received January 30, 2015; revised March 25, 2015.

K. L. Sam is with the Faculty of Social Sciences, Hong Kong Baptist University, Hong Kong (e-mail: samsam@life.hkbu.edu.hk).

C. X. Li and S. K. Lo are with the Faculty of Liberal Arts and Social Sciences, The Hong Kong Institute of Education, Hong Kong (e-mail: cxli@ied.edu.hk, skl@ied.edu.hk).

or remove from the whole scale. In reality, the unidimensional Rasch analysis is always be separately used to calibrate the single subscales. While the whole scale was in multidimensional construct, this approach, however, has both theoretical and practical limitations. For example, it does not take the correlation coefficient between latent traits into account, thence, might cause imprecise measures, especially when tests are short. In the theoretical aspect, this unidimensional assumption might not be appropriate for some psychological instruments that are designedly constructed from subcomponents which are contrived to measure multi traits [11]. Moreover, the demands of current psychometric assessment often fall beyond unidimensional analysis of person performances but require assessing the single trait from multiple facets [13], [14].

TABLE I: COMPARISONS OF CLASSICAL TEST THEORY (CTT), UNIDIMENSIONAL AND MULTIDIMENSIONAL RASCH MODELS

	Classical test theory	Uni-dimensional Rasch analysis	Multi-dimensional Rasch analysis
Data	Take ordinal scores as interval measure	Convert the ordinal data into interval measure	Convert the ordinal data into interval measure
Parameters	Sample dependence	Sample independence	Sample independence
Estimates	Within each subscale, all the items are estimated simultaneously	Calibrate separately for each subscale or take the total test as a whole by neglecting the nature of multidimensionality of the scale	Within each subscale, all the items are calibrated simultaneously
Correlations	Straightly estimate the correlations between factors based on the raw data	Underestimate the correlations due to measurement errors	The correlation between factors can be straightly estimated throughout the subscale calibrations

For the time being, multidimensional Rasch analysis has been widely utilized to validate the psychometric assessments, such as survey questionnaires in humanity or social sciences [14]-[19]. Except for some special multidimensional IRT models, e.g. testlet model or bifactor model, most multidimensional Rasch construct can be considered as a set of unidimensional subscales. So, as a good item, should solely measure a single trait to which it belongs to rather than other sub domains in the multidimensional construct. If the item set measures more than one dimension, including those with substantial differential item functioning (DIF), should be altered or substituted by others [11]. In the validation of psychological instruments, normally, the correlations between the latent traits are also investigated of great interests, considering as a multidimensional construct [20], [21]. The multidimensional model approach also preserves the subscale structure, calibrates all subscales simultaneously, and utilizes the correlations between subscales to increase measurement precision of each subscale. Comparing to the separate calibration for each subscale within a multiple-subscale test using unidimensional Rasch analysis, the multidimensional Rasch analysis can obtain a

much higher level of precise measure, and a more proper and accurate estimation for the correlations between subscales, particularly when tests are short and the number of tests are large [14], [15], [22]. A simple comparison of the classical testing method, unidimensional and multidimensional Rasch models is shown in Table I.

II. METHOD

A. Participants

Participants of the study was a convenient sample ($n = 521$) from ten universities located in the south-east of China. Five hundred and twenty one students (Male = 294, Female = 227) completed the study. Their age ranged from 18 to 25 years ($M = 21.3$, $SD = 1.38$). The majority participants were all in non-disability-related college majors (education technology, English language, sports science, and the related subjects among social sciences). Informed consent was obtained by their corresponding faculties/departments prior to the data collection.

B. Measures

The Chinese version of MRAI-R was obtained from Hampton and Xiao (2008) and was used to measure participants' attitudes toward the inclusion of people with ID. The MRAI-R consists of 29 items using a 4-point Likert scale (1 = Strongly Disagree, 4 = Strongly Agree) for responses. Three subscales (INSE, PRRT, and SUDB) have seven items and one (SDIS) consists of eight items. A total score of the MRAI-R is the sum of the responses and a higher score suggests more positive attitudes toward individuals with ID. Table II shows the subscales and items comprising the MRAI-R structure.

TABLE II: THE SUBSCALES AND ITEMS OF THE MRAI-R

Subscale	Number of items	Item number
Integration-Segregation (INSE)	7	1, 2, 7, 13, 17, 23, 29
Social Distance (SDIS)	8	3, 5, 11, 15, 18, 19, 24, 27
Private Rights (PRRT)	7	6, 8, 12, 14, 20, 22, 28
Subtle Derogatory Beliefs (SUDB)	7	4, 9, 10, 16, 21, 25, 26

In multidimensional Rasch models, the item bias can sufficiently be detected on the item level. To detect the item bias for students from different groups or subgroups, DIF is one of the analyses. If a scale is developed to work on different groups with equivalent level, such as male and female students, the items should not contain any bias in the scale for either group. DIF exists when an item does not correlate to the latent trait across different groups of students [23]. For an item without DIF, the item parameter estimate should equally be sufficient across various groups. In fact, several techniques have been used to detect the DIF, such as Mantel-Haenszel procedure [24] and the logistic regression procedure [25]. One of the widely used methods within the framework of IRT models is the likelihood ratio tests. In the tests, the logit difference between the constrained model with and without DIF was always adopted to examine DIF [26], [27]. For the MRAI-R, no attempts had been made to

scrutinize the psychometric properties on the item level including DIF. The gender and the subject group DIF on the item level had since not been detected.

The multidimensional random coefficients multinomial logit model, MRCMLM [13] was employed in this study. Under MRCMLM, the probability of a response in category k of item j for person i is

$$p_{ij}(\theta_n) = \frac{\exp(b_{ij}^T \theta_n + a_{ij}^T \xi)}{\sum_{u=1}^{K_i} \exp(b_{iu}^T \theta_n + a_{iu}^T \xi)} \quad (2)$$

where $\theta_n = (\theta_{n1}, \dots, \theta_{nD})$ is person n 's levels on the D latent traits, K_i is the number of categories in item i , ξ is a vector of location parameters that describe the items, b_{ij} is a score vector (known a priori) given to category j of item i across the D latent traits, and a_{ij} is a design vector given to category j of item i that describes the linear relationship among the elements of ξ . Eq. (2) can be expressed as below

$$\begin{aligned} \text{logit}_{ijn} &= \log\left(\frac{p_{ij}(\theta_n)}{p_{i(j-1)}(\theta_n)}\right) \\ &= (b_{ij}^T - b_{i(j-1)}^T)\theta_n + (a_{ij}^T - a_{i(j-1)}^T)\xi \quad (3) \\ &= b_{ij}^{*T}\theta_n + a_{ij}^{*T}\xi \end{aligned}$$

which is more consistent with the standard expression of the family of Rasch models. In Eq. (3), the item and person parameters are placed on the same logit scale [14], [15], [18].

A. Procedure

Permissions for administrating questionnaires were obtained from the research committee of investigators' institutions. Questionnaires were administrated by research assistants to those who agreed to join the study. A total of 600 questionnaires were distributed and 521 college students completed the questionnaire, with response rate of 86.83%.

B. Data Analysis

The partial credit model and the rating scale model were employed to fit the dataset respectively [12], [28]. The appropriate model with better goodness of fit was hence selected in the analysis below.

The DIF analysis was used for the detection purposes, based on two main demographic characteristics: gender (female and male) and major (education technology, English language, sports science, media, biology, physics, chemistry, public management, e-business, history, geography, tourism, and mathematics). The differences of the overall item difficulties between the female and male students and the maximum differences between major subject groups were calculated to detect the DIF effect.

The statistics software ConQuest [29] was utilized to do the item calibration and DIF analysis. This computer software incorporates the MRCMLM and has the advantage of calibrating multidimensional dataset with different class of Rasch models.

Before performing any statistical analysis, the items with negative meaning were reverse scored. The original MRAI-R scale was displayed in the Appendix.

III. RESULTS

A. Rating Scale Model versus Partial Credit Model

The multidimensional rating scale model, estimating 32 parameters, yielded a deviance ($-2\log\text{likelihood}$) of 28,622.598. In comparison with the multidimensional partial credit model, estimated 87 parameters and yielded a deviance of 27,905.136, the deviance of the multidimensional rating scale model was greater than that of the multidimensional partial credit model, and the chi-square test was shown extremely significant, which indicated that the partial credit model fit the dataset better ($\chi^2 = 717.462$, $df = 55$, $p < 0.001$).

B. Model-Data Fit

The second index to ascertain the goodness of fit of the model was the squared mean, MNSQ, value for the items and step parameters. In the Rasch analysis, item with the MNSQ value ranging from 0.70 to 1.30 is thought of a fitting item which implicates adequate construct homogeneity with other items in a scale to assure unidimensionality of the scale. Item with the MNSQ value outside the critical range is distinguished as possible misfitting items, indicating that further investigation may be warranted [30].

According to the analyses results, all the MNSQ values of the item and step parameters under the partial credit model were within the acceptable interval (ranged from 0.94 to 1.22). However, for the rating scale model, three of the items, item 20, 22, and 29, were out of the above-mentioned critical range (0.70 to 1.30). The MNSQ value of item 20 ("Campground and amusement park owners have the right to refuse to serve anyone they please, even if it means refusing people with ID") 22 ("If I were a barber or beauty shop owner I would not resent it I were told that I had to serve people with ID"), and 29 ("The child with ID should be integrated into regular classes in school") were 1.55, 1.80, and 1.49 respectively, which indicated that rating scale model did not fit the dataset. The above fit results indicated that the currently multidimensional partial credit model was more appropriate to be applied to the construct validity analysis and DIF analysis of the MRAI-R.

C. DIF Analysis

In the analysis of DIF, all of the differences were smaller than 0.10 logits except item 2 ("We should integrate people who have ID and who do not have ID into the same neighborhoods") and 8 ("Regardless of his or her own views, a private nursery school director should be required to admit children with ID"), which were 0.109 and 0.134 in the item parameters between genders, and, in advance, none of them were greater than 0.20 logits. Therefore, no items had substantial DIF.

D. Category Boundaries

The estimations of the item difficulties and step parameters were listed in Table III.

The step difficulty parameters for all the items within the four-factor structure ranged sufficiently large and were different for the items within or between the four subscales, which means that the persons took a different criterion for the four response categories for different items. For example, the three step difficulty parameters for item 1 ("School officials should not place children with ID and Children without ID in

the same classes”) were -2.694, 0.815, and 1.879, respectively, and the three step difficulty parameters for item 6 (“If I were a landlord, I would want to pick my tenants even if this meant only renting to people who do not have ID”) were -1.264, -0.812, and 2.075, which were significantly different from that of item 1.

On the other hand, the results above indicated that the order of the step difficulty estimates across the items was generally consistent. All of the items, the difficulties of the steps were ordered so that the later steps were more difficult than earlier steps, excepting item 2 (“We should integrate people who have ID and who do not have ID into the same neighborhoods”) and item 22 (“If I were a barber or beauty shop owner I would not resent it I were told that I had to serve people with ID”), in which the order of Step 1 and Step 2 difficulty estimates was reversed (Please see TableIII).

TABLE III: CATEGORY BOUNDARIES AND ITEM DIFFICULTY PARAMETERS OF MRAI-R

Item number	Difficulty	Step 1	Step 2	Step 3
Integration-segregation (INSE)				
1	0.563	-2.694	0.815	1.879
2	-1.225	-1.373	-1.394	2.768
7	0.081	-2.032	-0.205	2.237
13	0.131	-2.522	0.122	2.399
17	0.008	-2.098	-0.569	2.667
23	0.558	-2.734	0.326	2.408
29	-0.115	-2.296	0.152	2.144
Social Distance (SDIS)				
3	-0.103	n/a*	-2.013	2.013
5	-0.346	-1.898	-0.982	2.880
11	-0.082	-1.717	-1.198	2.915
15	-0.061	-2.895	-0.537	3.432
18	-0.009	-2.638	-0.537	3.175
19	0.422	-2.015	-1.499	3.514
24	0.198	-1.832	-1.495	3.327
27	-0.019	-2.231	-0.995	3.226
Private Rights (PRRT)				
6	0.059	-1.264	-0.812	2.075
8	-0.857	-2.946	0.059	2.887
12	0.092	-1.601	-0.658	2.259
14	0.375	-2.799	0.149	2.649
20	-0.538	-1.070	0.056	1.013
22	0.232	-0.232	-0.974	1.207
28	0.637	-2.630	0.442	2.188
Subtle Derogatory Beliefs (SUDB)				
4	0.496	-2.761	0.776	1.985
9	-0.264	-2.519	0.237	2.282
10	-0.195	-2.691	0.228	2.463
16	-1.262	-1.113	-0.944	2.057
21	0.023	-2.087	-0.176	2.263
25	0.816	-2.111	0.833	1.278
26	0.386	-2.616	0.514	2.102

* No information was received from the respondents in category 1 of item 3 (“I would allow my child to accept an invitation to a birthday party given for a child with ID”).

A. Correlations between Subscales

The correlations and covariances (with reliability row

below) between the subscales of MRAI-R were listed in Table IV.

According to the results, all the correlations between subscales ranged from 0.412 to 0.827.

B. Reliability Tests

The reliabilities for the subscales of MRAI-R deduced from multidimensional Rasch analysis were listed in the reliability row of Table IV. The results indicated that the reliabilities of the subscales except SUDB ranged from 0.664 to 0.833. The moderate to high reliabilities indicated good internal consistency of the items within each factor. However, the reliability of SUDB was only 0.496, which fell in the zone of low to moderate. This means that the factor SUDB had a relatively low internal consistency among the items. More investigation was needed to be done to improve the test reliability of the factor SUDB.

TABLE IV: RELIABILITIES, CORRELATIONS AND COVARIANCES BETWEEN THE SUBSCALES OF MRAI-R

	INSE	SDIS	PRRT	SUDB
INSE		0.496	0.264	0.165
SDIS	0.575		0.694	0.235
PRRT	0.704	0.827		0.127
SUDB	0.647	0.412	0.511	
Reliability	0.664	0.833	0.748	0.496

Values below the diagonal are correlations and values above with italics are covariances: INSE: Integration-segregation; SDIS: Social Distance; PRRT: Private Rights; SUDB: Subtle Derogatory Beliefs

IV. DISCUSSION

The present study aimed to explore the reliability and validity of the MRAI-R in Chinese college students through a multidimensional Rasch analysis. Compared with the CTT, the unidimensional IRT or Rasch analysis, and other methods, the multidimensional Rasch analysis can simultaneously take the inter-relationship between the subscales into account, as well as provide more detailed information at the item level [11]. Thus, the accuracy of correlations between the subscales and test reliability will be improved because the multidimensional Rasch analysis model estimates logits by considering the measurement error while conducting the item calibration.

The current research applied the MRCMLM [13] to investigate the construct validity of MRAI-R and employed DIF method to detect the item bias within the scale. Multidimensional Rasch analysis can provide more psychometric information on the item level as well as the construct validity.

The model-data fit indicated a good construct validity of the MRAI-R with the analysis results of the multidimensional partial credit model. In the results of step analysis (category boundaries), two items (item 2 and item 22) were disordered in their step levels. Although partial credit model does not require that the difficulties of the steps be ordered that the later steps are more difficult than earlier steps [31], the five categories labeling from “strongly disagree” to “strongly agree” represent a monotone increasing measure score on the scale of the latent traits which these items were measuring.

This implicates that some grouping of categories might be required for these items in the case of overlapping.

The results of DIF analysis suggested that all the items were met the assumption of unidimensionality when they were administered to different subject groups of students, and only two items (item 2 and item 8) exhibited little bias between genders.

According to the findings, the correlations between subscales (Table IV) were categorized as low-to-moderate, which might somewhat confirm the multidimensional structure of the full scale. However, the results also indicated that the subscale of SUDB had a relatively low reliability (0.496). By simple mathematics, if the subscale needed to increase to a certain level of reliability, the Spearman–Brown prophecy formula might be adopted to calculate the test length increment. The Spearman–Brown prophecy formula is shown as below

$$N = \frac{p_{xx}^*(1-p_{xx})}{p_{xx}(1-p_{xx}^*)} \quad (4)$$

where N is the number of item set, p_{xx} is the reliability of the current item set, and p_{xx}^* is the targeted reliability. For example, if the reliability of SUDB increases to 0.700, so called an “acceptable” level, according to the Spearman–Brown prophecy formula, the item set (test length) will be increased from 7 items to 16 or 17 items.

Besides, it cannot be denied that the item contents itself might play a crucial role on its validity and reliability. Although some indicators, i.e., face validity and construct validity, were reported in Hampton & Xiao (2008)’s study [7], the associated issues regarding to the consistency of the item set are still dubious. For the subscale of SUDB above, whether the test length or the item contents accounted for its low reliability, future work is needed to investigate the details of the item nature, in terms of content, so as to develop an appropriate (relatively reliable and valid) tool to somehow provide a holistic view on evaluating the current attitudes toward societal inclusion of people with ID.

This study evaluates the quality of MRAI-R from the view of statisticians; the technical terms, as well as the formula, used here are not intended for limiting the size of audience; on the other hand, some highlighted words, such as “testee” and “difficulty”, may not be entirely applicable for an attitude survey, since IRT approach itself was primarily devised for how to effectively differentiate testees’ ability from a tester’s perspective; as for evaluation purposes, it must be cautious when interpreting those “scores”; further elaboration and clarification are deemed important always.

V. CONCLUSION

MRAI-R is commonly used for investigating people’s attitudes toward ID. Rasch analysis in the current study revealed that the MRAI-R exhibits an adequate validity and reliability for the Chinese college students, which is coherent to the similar studies in US and Japan [8-10]. Further examination in item nature is needed so as to develop a relatively robust tool for the investigation of attitudes on societal inclusion of people with ID.

APPENDIX

Mental Retardation Attitude Inventory-Revised (MRAI-R):

Integration-segregation (INSE)	
1.	School officials should not place children with ID and Children without ID in the same classes
2.	We should integrate people who have ID and who do not have ID into the same neighbourhoods
7.	It is a good idea to have separate after-school programmes for children who have ID and children who do not have ID
13.	Integrating children who have ID and who do not have ID into the same pre-school classes should not be attempted because of the turmoil it would cause
17.	Having people who have ID and who do not have ID work at the same jobsites will be beneficial to both
23.	Assigning high school students who have ID and who do not have ID to the same classes is more trouble than it is worth
29.	The child with ID should be integrated into regular classes in school
Social Distance (SDIS)	
3.	I would allow my child to accept an invitation to a birthday party given for a child with ID
5.	I am willing for my child to have children who have ID as close personal friends
11.	I have no objection to attending the movies or a play in the company of people with ID
15.	I would rather not have people with ID as dinner guests with my friends who are not ID
18.	I would rather not have a person who has ID swim in the same pool that I swim in
19.	I would be willing to introduce a person with ID to friends and neighbours in my home town
24.	I would be willing to go to a competent barber or hairdresser with ID
27.	I would rather not have people with ID live in the same apartment building I live in
Private Rights (PRRT)	
6.	If I were a landlord, I would want to pick my tenants even if this meant only renting to people who do not have ID
8.	Regardless of his or her own views, a private nursery school director should be required to admit children with ID
12.	Laws requiring employers not to discriminate against people with ID violate the rights of the individual who does not want to associate with people who are ID
14.	Real estate agents should be required to show homes to families with children who have ID regardless of the desires of the homeowners
20.	Campground and amusement park owners have the right to refuse to serve anyone they please, even if it means refusing people with ID
22.	If I were a barber or beauty shop owner I would not resent it I were told that I had to serve people with ID
28.	A person should not be permitted to run a day care centre if he or she will not serve children who have ID
Subtle Derogatory Beliefs (SUDB)	
4.	People who have intellectual disability are not yet ready to practice the self-control that goes with social equality with people without ID
9.	Even though children with ID are in public school, it is doubtful whether they will gain much from it
10.	Although social mixing of people who have ID and who do not have ID may be right, it is impractical until people with ID learn to accept limits in their relations with the opposite sex
16.	Children who are ID waste time playing in class instead of trying to do better
21.	The problem of prejudice towards people with ID has been exaggerated
25.	Even with equality of social opportunity, people with ID could not show themselves equal in social situations to people without ID
26.	Even though people with ID have some cause for complaint, they would get what they want if they were more patient

The Chinese version of MRAI-R was obtained from Hampton and Xiao (2008), and the original English version of MRAI-R was developed by Antonak and Harth (1994)

REFERENCES

- [1] C. Bigby, "Shifting models of welfare: Issues in relocation from an institution and the organization of community living," *Journal of Policy and Practice in Intellectual Disabilities*, vol. 3, no. 3, pp. 147-154, 2006.
- [2] H. O. Kuntz *et al.*, "Public attitudes towards individuals with intellectual disabilities as measured by the concept of social distance," *Journal of Applied Research in Intellectual Disabilities*, vol. 23, no. 2, pp. 132-142, 2010.
- [3] R. A. Cummins and A. L. D. Lau, "Community integration or community exposure? A review and discussion in relation to people with disability," *Journal of Applied Research in Intellectual Disabilities*, vol. 16, no. 2, pp. 145-157, 2003.
- [4] F. Xiao, "Mainstreaming in China: History, actuality, perspectives," *Chinese Journal of Special Education*, vol. 3, pp. 3-7, 2005.
- [5] N. Z. Hampton and F. Xiao, "Traditional Chinese values and attitudes of Chinese university students toward people with intellectual disabilities," *International Journal of Disability, Development and Education*, vol. 56, no. 3, pp. 247-261, 2009.
- [6] N. Z. Hampton and F. Xiao, "Attitudes toward people with developmental disabilities in Chinese and American students: the role of cultural values, contact, and knowledge," *Journal of Rehabilitation*, vol. 73, no. 4, pp. 23-32, 2007.
- [7] N. Z. Hampton and F. Xiao, "Psychometric properties of the mental retardation attitude inventory-revised in Chinese college students," *Journal of Intellectual Disability Research*, vol. 52, no. 4, pp. 299-308, 2008.
- [8] R. F. Antonak and R. Harth, "Psychometric analysis and revision of the Mental Retardation Attitude inventory," *Mental Retardation*, vol. 32, no. 4, pp. 272-280, 1994.
- [9] J. Krajewski and T. Flaherty, "Attitudes of high school students toward individuals with mental retardation," *Mental Retardation*, vol. 38, no. 2, pp. 154-162, 2000.
- [10] W. H. Johnson *et al.*, "Attitudes of Japanese students toward people with intellectual disability," *Journal of Intellectual Disability Research*, vol. 46, no. 5, pp. 365-378, 2002.
- [11] J. Lee, Z. Zhang, and H. Yin, "Using multidimensional Rasch analysis to validate the Chinese version of the Motivated Strategies for Learning Questionnaire (MSLQ-CV)," *European Journal of Psychology of Education*, vol. 25, no. 1, pp. 141-155, 2010.
- [12] G. Masters, "A Rasch model for partial credit scoring," *Psychometrika*, vol. 47, no. 2, pp. 149-174, 1982.
- [13] R. J. Adams, M. R. Wilson, and W. C. Wang, "The multidimensional random coefficients multinomial logit model," *Applied Psychological Measurement*, vol. 21, no. 1, pp. 1-23, 1997.
- [14] W. C. Wang, P. H. Chen, and Y. Y. Cheng, "Improving measurement precision of test batteries using multidimensional item response models," *Psychological Methods*, vol. 9, no. 1, pp. 116-136, 2004.
- [15] Y. Y. Cheng, W. C. Wang, and Y. H. Ho, "Multidimensional Rasch analysis of a psychological test with multiple subtests: A statistical solution for the bandwidth-fidelity dilemma," *Educational and Psychological Measurement*, vol. 69, no. 3, pp. 369-388, 2009.
- [16] T. W. Chien *et al.*, "Using Rasch analysis to validate the revised PSQI to assess sleep disorders in Taiwan's hi-tech workers," *Community Mental Health Journal*, vol. 44, no. 6, pp. 417-425, 2008.
- [17] K. Pesudovs and B. A. Noble, "Improving subjective scaling of pain using Rasch analysis," *The Journal of Pain*, vol. 6, no. 9, pp. 630-636, 2005.
- [18] W. C. Wang *et al.*, "Validating, improving reliability, and estimating correlation of the four subscales in the WHOQOL-BREF using multidimensional Rasch analysis," *Quality of Life Research*, vol. 15, no. 4, pp. 607-620, 2006.
- [19] L. Yao and R. D. Schwarz, "A multidimensional partial credit model with associated item and test statistics: an application to mixed-format tests," *Applied Psychological Measurement*, vol. 30, no. 6, pp. 469-492, 2006.
- [20] I. Koller and C. Lamm, "Item response model investigation of the (German) interpersonal reactivity index empathy questionnaire: Implications for analyses of group differences," *European Journal of Psychological Assessment*, No Pagination Specified, 2014.
- [21] S. M. van den Berg, M. C. S. Paap, and E. M. Derks, "Using multidimensional modeling to combine self-report symptoms with clinical judgment of schizotypy," *Psychiatry Research*, vol. 206, no. 1, pp. 75-80, 2013.
- [22] L. Yao and K. A. Boughton, "A multidimensional item response modeling approach for improving subscale proficiency estimation and classification," *Applied Psychological Measurement*, vol. 31, no. 2, pp. 83-105, 2007.
- [23] S. E. Embretson and S. P. Reise, *Item response Theory for Psychologists*, L. Erlbaum Associates, 2000.
- [24] P. W. Holland and D. T. Thayer, "Differential item performance and the Mantel-Haenszel procedure," in *Test validity*, H. Wainer and H. I. Braun Ed., Erlbaum: Hillsdale, 1988, pp. 129-145.
- [25] H. J. Rogers and H. Swaminathan, "A comparison of logistic regression and Mantel-Haenszel procedures for detecting differential item functioning," *Applied Psychological Measurement*, vol. 17, no. 2, pp. 105-116, 1993.
- [26] A. S. Cohen, S. H. Kim, and J. A. Wollack, "An investigation of the likelihood ratio test for detection of differential item functioning," *Applied Psychological Measurement*, vol. 20, no. 1, pp. 15-26, 1996.
- [27] W. C. Wang and M. Wilson, "Assessment of differential item functioning in testlet-based items using the Rasch testlet model," *Educational and Psychological Measurement*, vol. 65, no. 4, pp. 549-576, 2005.
- [28] D. Andrich, "A rating scale formulation for ordered response categories," *Psychometrika*, vol. 43, no. 4, pp. 561-573, 1978.
- [29] M. L. Wu, R. J. Adams, and M. R. Wilson, "Conquest," *Camberwell: Australian Council for Educational Research*, 1998.
- [30] B. D. Wright *et al.*, "Reasonable mean-square fit values," *Rasch Measurement Transactions*, vol. 8, no. 3, pp. 370, 1994.
- [31] B. G. Dodd and W. R. Koch, "Effects of variations in item step values on item and test information in the partial credit model," *Applied Psychological Measurement*, vol. 11, no. 4, pp. 371-384, 1987.



Sam Ka-Lam holds a doctor of education, degree of master of education, and degree of master of science.

Dr. Sam obtained his EdD degree from the Hong Kong Institute of Education (HKIEd). He is an experienced educator, and is actively conducting research from psychoeducational perspectives. His areas of research mainly include advancement of statistical models, development of psychometric tools, and school-based intervention for children with special educational needs.



Li Chunxiao holds a doctor of philosophy, degree of master of education, and degree of bachelor of education.

Dr. Li obtained his PhD degree from National Institute of Education, Nanyang Technological University. He is currently a lecturer of the Department of Health and Physical Education, HKIEd. He teaches measurement & evaluation, statistics, research methods, history & philosophy, and various sport modules. His areas of research mainly include psychosocial aspects of talent development, social inclusion, moral behaviors, scale development, systematic reviews, as well as training and athlete motivation.



Lo Sing Kai holds a doctor of philosophy. Prof. Lo is currently the chair professor of the Faculty of Liberal Arts and Social Sciences, HKIEd. He has a solid track record in securing research incomes; and has published over 200 articles, mainly in journals with high impact factors. He has also won numerous teaching awards. He has a strong background in capacity building and is particularly experienced in providing mentorship services to early and mid-career researchers.