Northumbria Research Link

Assessing medical mistrust in organ donation across countries using item response theory

Arunangshu Ghoshal, Ronan O'Carroll, Eamonn Ferguson, Lee Shepherd, Sally Doherty, Mary Mathew,

Karen Morgan, Frank Doyle

Authors:

1. Arunangshu Ghoshal, MD, MRes

a. Dept. of Palliative Medicine, Tata Memorial Hospital, Tata Memorial Center, Mumbai, India

b. Homi Bhaba National Institute (HBNI), Mumbai, India. E: arunghoshal@outlook.com

2. Ronan E. O'Carroll, PhD

Division of Psychology, University of Stirling, Stirling, UK. E: ronan.ocarroll@stir.ac.uk

3. Eamonn Ferguson, PhD

School of Psychology, University of Nottingham, Nottingham, UK. E:

Eamonn.Ferguson@nottingham.ac.uk

4. Lee Shepherd, PhD

Department of Psychology, Northumbria University, Newcastle upon Tyne, UK. E:

lee.shepherd@northumbria.ac.uk

5. Dr Sally Doherty PhD, MSc, PGCert Higher Ed, CPsychol

Senior Lecturer in Psychology, RCSI Bahrain, Adliya, Kingdom of Bahrain. E: sdoherty@rcsi-

mub.com

6. Mary Mathew MD, PGDMLE

Department of Pathology, Kasturba Medical College, Manipal Academy of Higher Education,

Karnataka, Manipal, India. E: marymathew883@gmail.com

7. Karen Morgan PhD, C. Psychol., Ps.S.I, PGCert Med Ed, FEHPS

Deputy Dean, Perdana University-Royal College of Surgeons in Ireland School of Medicine

DU 026(B), Suite 9.2, 9th Floor, Wisma Chase Perdana, Changkat Semantan, Damansara

Heights, Kuala Lumpur, Malaysia. E: karenmorgan@perdanauniversity.edu.my

8. Frank Doyle F. Ps.S.I

Senior Lecturer, Division of Population Health Sciences, Lower Mercer Street, Dublin, Ireland.

E: FDoyle4@rcsi.ie

Corresponding Author

Arunangshu Ghoshal, MD, MRes

Dept. of Palliative Medicine, Tata Memorial Hospital, Tata Memorial Center, Homi Bhaba National

Institute (HBNI)

Dr E Borges Road, Mumbai 400012, India. T: +918240675031 E: arunghoshal@outlook.com

Running Head: Assessing medical mistrust using item response theory

Declarations

Compliance with Ethical Standards

Disclosure of potential conflicts of interest – The authors have no conflicts of interest to declare that are relevant to the content of this article.

Research involving human participants and/or animals – All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The protocol was approved by the Research Ethics Committees (RECs) in each site (RCSI Dublin REC1048bb; UK—Stirling University General University Ethics Panel [No. 188]; Malaysia, RCSI PU REC—PUIRBH0097; India—Kasturba Hospital Institutional Ethics Committee IEC 134/2017).

Informed consent – Informed consent was obtained from all individual participants included in the study.

Data, Materials and/or Code availability – The current article includes the complete raw dataset collected in the study including the participants' data set, syntax file and log files for analysis. Pending acceptance for publication, all of the data files will be automatically uploaded to the Figshare repository.

Authors' contribution statements –

1) made substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data; or the creation of new software used in the work: AG, REOC, SD, MM, KM, FD

2) drafted the work or revised it critically for important intellectual content: AG, FD

3) approved the version to be published: Arunangshu Ghoshal (AG), Ronan E. O'Carroll (REOC), Eamonn Ferguson (EF), Lee Shepherd (LS), Sally Doherty (SD), Mary Mathew (MM), Karen Morgan (KM), Frank Doyle (FD)

4) agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved: AG, FD

Word Count: abstract 140, main text: 4072

1   **Assessing medical mistrust in organ donation across countries using item response theory**

2   **Abstract**

3   Although medical mistrust (MM) may be an impediment to public health interventions, no MM

4   scale has been validated across countries and the assessment of MM has not been explored using

5   item response theory, which allows generalization beyond the sampled data. We aimed to

6   determine the dimensionality of a brief MM measure across four countries through Mokken

7   analysis and Graded Response Modelling. Analysis of 1468 participants from UK (n=1179),

8   Ireland (n=191), India (n=49) and Malaysia (n=49) demonstrated that MM items formed a

9   hierarchical, unidimensional measure, which is very informative about high levels of MM.

10   Possible item reduction and scoring changes were also demonstrated. This study demonstrates

11   that this brief MM measure is suitable for international studies as it is unidimensional across

12   countries, cross cultural, and shows that minor adjustments will not impact on the assessment of

13   MM when using these items.

14   **Keywords**

15   medical mistrust; organ donation; Mokken analysis; Item response theory.

16

1 **Main text**

2 **INTRODUCTION**

3 Medical mistrust (MM) is a tendency to distrust medical systems and personnel believed to

4 represent the dominant culture in a society. Medical "distrust" and "mistrust" are different terms,

5 but sometimes are used interchangeably in literature – "distrust" is most correctly based on a

6 specific experience or information, whereas, "mistrust" expresses a general sense of unease

7 toward someone or something (Smirnoff et al., 2018). MM is an active response to direct or

8 vicarious experiences of marginalization or poor experiences, including medical error (Williams

9 and Mohammed, 2009). For example, having a higher mistrust, a person receiving a cancer

10 diagnosis may doubt the doctor's advice and seek a second opinion(Guadagnolo et al., 2009), or

11 a family member with high mistrust may be less likely to consent to organ donation when it is

12 raised(Morgan, Harrison, et al., 2008), or a person with high mistrust may be less likely to seek

13 medical attention when experiencing symptoms that they can't explain(Taber et al., 2015).

14 MM is often context specific and has significant negative public health implications, like lower

15 adherence to cancer screening guidelines (Adams et al., 2017), lower adherence to antiretroviral

16 medication among men with HIV (Galvan et al., 2017), poor patient-provider relationships

17 (Saunders, 2017), engaging in risky behaviours (Hardin et al., 2018), conspiracy theories (Quinn

18 et al., 2018), and lower use of vaccinations (Benkert et al., 2019; Larson et al., 2018), and even

19 the recent resurgence of anti-vaccination movement in Western countries (Hussain et al., 2018;

20 Opel et al., 2021).

21 Practicing organ donation has proven to be lifesaving, but mMistrust in the healthcare system is

22 also associated with reduced intentions for organ donation (Miller et al., 2019; Russell et al.,

23 2012). In the US alone, 64.02% of candidates failed to receive a transplant in 2020 (Detailed

Description of Data | organdonor.gov, n.d.), highlighting ~~Indeed, anti-vaccination has been listed~~ ~~as one of the top ten threats to global health (Aranda, 2019), highlighting~~ the importance of appropriate assessment of MM ~~internationally and across cultures. Mistrust~~ in the healthcare ~~system is also associated with reduced intentions for organ donation (Miller et al., 2019; Russell~~ ~~et al., 2012), and is the area of interest of this paper.~~

However, few studies have attempted an in-depth exploration of the assessment of MM. In a recent systematic review (Ozawa and Sripad, 2013), around 20% studies used measures with 2-3 items or single item, such as "It is difficult for me to trust doctors and other health professionals" (Williamson and Bigman, 2018). Commonly-used, multi-item scale measures include: the 12-item Group-Based Medical Mistrust Scale (Thompson et al., 2004), which was developed to measure race-based medical mistrust; the 7-item Medical Mistrust Index (Sheppard et al., 2019), which assesses the association between mistrust and five measures of underuse of health services; and the 10-item Health Care System Distrust Scale (Rose et al., 2004).

Previous psychometric work has identified three subscales within the Group-based Medical Mistrust Scale: (1) Lack of Support from doctors and health care workers; (2) Discrimination and group disparities in health care; and (3) Suspicion of doctors, health care workers, and medicine (Shelton et al., 2010). There are important differences among these scales with regards to the object of mistrust (e.g., system, individual physician), referent specificity (e.g., group), health topic, and sample population in which they were developed, with around 88% being US-based. In addition, most were not used internationally nor validated across countries (Williamson and Bigman, 2018) though research shows that mistrust varies by ethnicity and within countries (Henrich et al., 2012). To establish psychometric validity, the Medical Mistrust Index and the Health Care System Distrust Scale relied on Classical Test

Theory (CTT) parameters (DeVellis, 2006), reporting Cronbach's alpha reliability for scale

performance, without first adopting the recommended factor analytic techniques to establish

dimensionality (Crutzen and Peters, 2017). In addition, while the Group-Based Medical Mistrust

Scale used exploratory factor analysis (Gaskin and Happell, 2014), and structural equation

modelling (SEM) (Tarka, 2018) in its development process, these methods can be criticised for

being sample-dependent (Hays et al., 2000) – so it is unknown if the findings generalise across

countries and cultures. To overcome limitations such as these, a psychometric technique

increasingly used is the Item Response Theory (IRT) approach – a family of mathematical

models that attempt to explain the relationship between latent traits (unobservable characteristic

or attribute) and their manifestations (i.e. observed outcomes, responses or performance) (Hays

et al., 2000). IRT has several advantages over CTT, including that rather than looking at the

reliability of the test as a whole, IRT looks at each item individually. IRT suggests that a

respondent's latent ~~ability~~ trait is independent from the item or question on a test (Jabrayilov et

al., 2016). It produces person-parameter invariance (i.e. test scores are not dependent on the

particular choice of test items) when the model fit is present, while test information functions

provide the amount of information or "measurement precision" captured by the test on the scale

measuring the construct of interest (Petrillo et al., 2015). IRT can, therefore, provide information

for scale development and scoring that is not sample-dependent, is generalisable across cultures,

countries, and contexts that is not readily available via CTT methods, and arguably robust

psychometric scale evaluation is incomplete without such assessments. Individuals might have

different opinions about MM, but IRT methodology ensures that the results do not vary with

population characteristics (Hambleton, 2009). Rasch analysis is a special case of IRT, which

requires the data to fit the model in order to generate invariant, interval-level measures of items

1 and persons, but is prescriptive by nature and assumes that each item has the same discrimination

2 level (Boone et al., 2014). Rasch analysis was used to determine the validity of an instrument for

3 measuring patients' trust in medical technology (Montague, 2010), but to date, no studies have

4 assessed MM using more flexible and powerful IRT analyses.

5 To summarise, there is limited psychometric evaluation of MM scales, including a lack of

6 exploration across countries and cultures which limits comparability and generalisability, and no

7 study to date has investigated MM using IRT. In addition, with challenges such as the current

8 global threat of the anti-vaccination movement and the need for organs for transplant, robust

9 measures are urgently needed to better understand potential contributors to MM. We therefore

10 aimed to fill this gap in the literature, by analysing MM items that were previously developed by

11 Morgan, Stephenson, Harrison, Afifi, & Long (2008) and used in several studies of MM in the

12 context of attitudes to organ donation (Doherty et al., 2017; Doyle et al., 2019; O'Carroll et al.,

13 2016). Specifically, we applied nonparametric and parametric IRT to investigate the

14 dimensionality of MM across four countries. We explored whether the MM latent trait is

15 adequately assessed by the MM items developed by Morgan et al., andet al. and investigated the

16 MM items and the scoring of these items, with a view to determining whether item reduction or

17 scoring changes would impact on MM assessment.

18 **METHODS**

19 Participants and procedure

20 We used combined data from four randomised experiments on attitudes to organ donation, for

21 secondary analysis of medical mistrust (Doherty et al., 2017; Doyle et al., 2019; O'Carroll et al.,

22 2016). In brief, these studies ascertained the effects of affective attitudinal items on intention and

registration for organ donation. The sample consisted of participating adults recruited from the community, shopping centres, libraries, campuses, and online from 4 countries — United Kingdom (313 in-person, 616 online, 864 postal), Ireland (586), Malaysia (148) and India (175). All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The protocol was approved by the Research Ethics Committees (RECs) in each site (RCSI Dublin REC1048bb; UK—Stirling University General University Ethics Panel [No. 188]; Malaysia, RCSI PU REC—PUIRBH0097; India—Kasturba Hospital Institutional Ethics Committee IEC 134/2017).

Measures

As part of the above studies, participants completed 4 items measuring medical mistrust, adapted from Morgan et al. (2008). Items were rated on a seven-point scale (1 =strongly disagree; 7 = strongly agree) and evidenced good reliability (α = .84). Items included were as follows: MM1: 'Hospitals sometimes prescribe medication as a way of experimenting on people without their knowledge or consent'; MM2: 'If I sign an organ donor card, doctors might take my organs before I'm actually dead'; MM3'Sometimes, medical procedures are done on people without their consent'; and MM4: 'If I sign an organ donor card, doctors might not try so hard to save my life'.

Statistics

We used non-parametric and parametric Item Response Theory (IRT) analysis. Mokken scaling is a non-parametric, iterative scale-building technique, especially suitable for skewed and polytomous items (van Schuur, 2003), which is ideal for establishing unidimensionality (Dima,

2018). Mokken scaling is a stochastic extension of Guttman scales (Rothman, 2004), which are unidimensional, ordinal scales of binary items along a continuum. A positive answer to one item of a certain 'difficulty' indicates that all other items of lesser difficulty have also been answered positively. Mokken scaling facilitates analyses of polytomous items with probabilistic ordering. Mokken scaling has several advantages over the typically adopted factor analytic techniques: it is suitable for binary or highly skewed polytomous items, it can be used in relatively small samples (n = 300-400), and it systematically orders items (Watson et al., 2012). Loevinger's H-coefficient is used for interpretation. According to (Sijtsma and Molenaar, 2011), $0.3 \geq H < 0.4$, $0.4 \geq H < 0.5$, and $H \geq 0.5$ indicate weak, moderate, and strong scales, respectively. Higher H values indicate higher item discrimination power, and thus more confidence in the ordering of respondents. Although this project is a secondary analysis, so sample size was not considered a priori, tThe total sample size far exceeds the required estimates to establish scalability (Watson et al., 2018), and therefore was used to assess the unidimensionality assumption of parametric IRT employed in the next steps of validation. Mokken scaling is an established method for IRT analysis with wide application in the social sciences. It provides psychometricians with an additional tool in the development of questionnaires and in the study of individuals and their responses to latent traits (Watson et al., 2012).

Model fit for parametric IRT (see appendix) was considered using the likelihood ratio test of Akaike and Bayesian information criteria (Swaminathan et al., 2006), which indicated that the unconstrained Graded Response Model (GRM) was most appropriate. The IRT GRM approach constructs a nonlinear relationship between latent traits and item responses (Samejima, 2016). The GRM is especially suited to ordinal response items, as it does not assume that each item discriminates equivalently along with the latent trait (unlike Rash models) (Pallant and Tennant,

2007). The GRM characterizes item functioning via two types of item parameters for each item: the item discrimination and item difficulty parameters. We report test-level and item-level parameters. The **test information function**, a function of the true score theta (θ, in this case, MM), is a measure of the amount of information provided by the item responses about MM. The **test characteristic curve** is the functional relation between the true score and the ~~ability~~ trait scale. **Item Characteristic Curve** shows the behaviour of individual items on a scale relative to the latent trait. Steeper curves indicate higher discrimination. **Item information function** is a measure of how much statistical information a test item provides. Item information is a function of MM. **Category characteristic curves** were used for the polytomous response categories of MM items. Each CCC depicts the probability of endorsing a single response category (Fayers, 2004; Ockey, 2013; Reckase, 2009).

Mokken scale analysis was conducted using a procedure written for Stata (StataCorp, 2019), by Jean-Benoit Hardouin (Hardouin, 2004), GRM IRT was conducted in Stata SE 15 (StataCorp, 2017). Other analyses were done in RStudio Version 1.3.959 (The R Foundation, 2018), using the procedure described by van der (Ark, 2007).

Data Sharing Statement: The current article includes the complete raw dataset collected in the study including the participants' data set, syntax file and log files for analysis. Pending acceptance for publication, all of the data files will be automatically uploaded to the Figshare repository.

RESULTS

Sample profile

1   The baseline profile of the larger samples has been described in detail elsewhere (Doherty et al.,

2   2017; Doyle et al., 2019; O'Carroll et al., 2016). The mean age for the total sample (n=2702) was

3   42.3 years (SD = 16.9), with 58% being women. The number of subjects who completed all MM

4   items was 1468 (54% response rate; due to the differing exposures across the experimental

5   groups) (mean age = 44 years, 60% were women). The details of responses as obtained from 4

6   countries were: United Kingdom (107 in-person [mean age = 46 years, 42% were women], 211

7   online [mean age = 51 years, 51% were women], 861 postal [mean age = 45 years, 66% were

8   women]), Ireland (191; (mean age = 37 years, 48% were women), Malaysia (49; mean age = 19

9   years, 55% were women) and India (49; mean age = 24 years, 81% were women).

10 *Mokken scaling results*

11 The overall scale H score of 0.51 (for individual items = 0.47 - 0.54) indicates a 'strong' scale,

12 showing that all items in this scale tapped a hierarchical, unidimensional variable of medical

13 mistrust measure (Table 1). We also found unidimensionality for each country (Ireland: H=.51,

14 UK (in person): H=.53 UK (online): H=.61, UK (postal): H=.54, Malaysia: H=.56, India: H=.48

15 (3 items). It should be noted that one item was dropped from the analysis for India as because the

16 item was not selected by the Mokken scaling procedure as it did not discriminate along the latent

17 traitit did not appear to discriminate along the MM trait: 'If I sign an organ donor card, doctors

18 might not try so hard to save my life'.

19 *Graded Response Model results*

20 Test level parameters

21 The Test Information Function for the overall assessment of the MM latent trait is shown in

22 Figure 1.1. Most information is provided at mid-higher levels of the latent trait, but very little

1   about lower levels of MM. As shown in Figure 1.2, the Test Characteristic Curve, there was a

2   substantial spread of scores evident, as here we expect 95% of people to score between 5 to 23.

3   <u>Item level parameters</u>

4   The Item Information Curves (Figure 2.1) show how well and precisely each item measures the

5   latent trait at various levels of the attribute. Items 1 and 3 provide more information at low levels

6   of the trait (MM) and across a broader range of scores, while items 2 and 4, which specifically

7   refer to organ donation, provide more information overall. As shown in the Item Characteristic

8   Curve (Figure 2.2), items 2 and 4, which concern mistrust in doctors, show significantly higher

9   levels of discrimination (95% CIs do not overlap), with the discrimination indices available in

10   Table 2. Also clearly indicated is that the discrimination parameters for scores of 3 and 4 within

11   individual items lie very close together, suggesting that, if necessary, these could be combined to

12   form an overall 1-6 scoring range without losing much information. (Also see correlation matrix

13   in appendix 2).

14   Category characteristic curves for MM items (Figure 3) represents the probability that a

15   respondent will select a response option, given the respondent's latent trait (MM) value. It

16   displays the probability of a person endorsing a particular response category based on their level

17   of support for the item and the intensity or difficulty of the item. The range of latent trait (MM)

18   value is plotted along the x-axis. The probability of selecting each category is plotted along the

19   y-axis. An item is better at discriminating between individuals when the curves are peaked and

20   dispersed across all levels of the latent trait. For example, an item with high discrimination

21   would have 7 peaks dispersed from low levels of the latent trait to high levels of the latent trait.

22   Here, there is an overall well-spaced response category across MM1-4, with MM4 having well

23   dispersed peak signifying better discrimination between individual scorings.

DISCUSSION

To our knowledge, this is the first study to apply flexible IRT analyses to the measurement of MM, which highlights a robust measure that is generalisable across cultures and countries that is preferable to scales developed in single-country settings. Although the items here relate to organ donation, and the test has implications for the organ donation context, the test is measuring something more than just attitudes towards organ donation, but towards medical professionals and medical treatment in general. We used Mokken scaling to demonstrate that the 4 MM items developed by Morgan et al. (Morgan, Stephenson, et al., 2008) formed a unidimensional scale across 4 countries. The parametric IRT GRM approach demonstrated that the measure assesses medium to higher levels of MM very well, but unsurprisingly provides less information at lower levels of MM. Furthermore, it was clear that items relating to doctors are significantly more highly discriminating MM than other items, which is a novel addition to the literature.

Our results support the original findings of Morgan et al., who used structural equation modelling on data from 4426 participants at six different geographic locations in the United States, and showed that MM was a unidimensional scale and one of the most influential noncognitive beliefs of an individual having a direct influence on organ donor card status (Morgan, Stephenson, et al., 2008). In addition, IRT analysis has enhanced the validity of the scale across 4 different countries and cultures and has highlighted unique aspects of overall scale and item level parameters. Incidentally, the scale with 4 items did not show validity for India, and one item ('If I sign an organ donor card, doctors might not try so hard to save my life') had to be dropped. This is probably due to lower sample size in this population, which does not meet the criteria for obtaining a valid overall scale H-value (Watson et al., 2012). It could also be due to high level of awareness of organ donation protocols in the same geographical region as has

been shown in a study by (Mithra et al., 2013), potentially introducing nonresponse bias (Sax et al., 2003), or due to high proportion of women responders as Data from the National Organ Transplantation and Tissue Organization show that majority of living donations in India are from women and majority of organ recipients are men (National Organ & Tissue Transplant Organization, n.d.)(Sahay, 2019). Further research is needed to explore this aspect in other parts of India as has been acknowledged elsewhere (Vincent et al., 2019).

The results of our IRT GRM-based analyses provide a closer look at the properties of the MM scale. The test and item information functions represent the precision with which a respondent's MM value can be estimated. Higher information indicates greater precision. The Test Information Function revealed that the measure provides most information (estimation precision) at moderate/high levels of MM, but little information at low levels. Similarly, the Item Information Function curves showed that items 1 and 3 provided more information at low levels of MM, while items 2 and 4 provide more information at higher levels. These results suggest that, if it were necessary to shorten the scale (for example when competing with other measures in large population-based surveys), researchers could base their decision on whether discrimination or latent trait coverage were considered more advantageous, along with, of course, the content validity of the items. This has been done elsewhere, for example when measuring quality of life in breast cancer survivors (Xia et al., 2019). However, the final decision about the exclusion of items should be based on substantive reasons and not merely through the acceptance of quantitative values as candidate threshold levels, including theory and content validity. This is especially important as we may lose information about part of the measured domain that might be useful or important otherwise (Edelen and Reeve, 2007).

1   The properties of the scale are depicted through the Category characteristic curves, which

2   shows overall well-spaced response categories across MM1-4. As shown in the item

3   characteristic curves was that the discrimination parameters for scores of 3 and 4 lie close

4   together, suggesting that these could be combined to form an overall 1-6 scoring range, without

5   losing precision as has been shown elsewhere (Chen et al., 2009; Edelen and Reeve, 2007).

6   However, caution should be used when making such decisions, as the trade-off for a reduced

7   number of response options may lead to a decline in scale precision, particularly for respondents

8   at extremes of the latent trait (Yang and Kao, 2014). Item reduction is a pragmatic choice in most

9   cases and should be individualized to the study as has been explained elsewhere (Kolva et al.,

10   2017).

11   In our findings, items which concerned mistrust in doctors showed significantly higher

12   levels of discrimination than items which referred to more systemic aspects, such as hospitals

13   and procedures. This is in contrast to other research, which shows that while distrust in the health

14   care system is relatively high (20% - 80% of respondents reporting distrust for each item on the

15   Health Care System Distrust scale) (Rose et al., 2004), there is in contrast lower mistrust in one's

16   own primary care physician (only 10% to 20% of respondents endorsing mistrust). This may be

17   due to the wording of the items – the questions in our study concerned doctors in general, not

18   specifically participants own doctors – future research could investigate the impact of such

19   wording changes on this MM measure.

20   The MM measure provides a substantial amount of information about high levels

21   mistrust, but little information around lower levels of the MM trait. We should not assume that

22   low levels of MM correlates directly with higher levels of trust – these factors may be

23   multidimensional, rather than unidimensional, as has been elucidated elsewhere (Goold, 2002;

Pellowski et al., 2017; Sullivan, 2020). To explore trust, it may be necessary to employ another measure, perhaps also including reference to one's own doctors rather than physicians in general, as outlined above. Trust is an important component of healthcare, and is elucidated as a set of expectations that the healthcare provider will do the best for the patient and has different meanings attached to it (Ozawa and Sripad, 2013). Thus, conducting cognitive interviews that focus on the wording or sense of these items may provide insight into respondents' interpretations and responses. Future research could also explore newly written items to determine if they indicate discrimination at the lower levels of the MM trait. Also, of interest would be conducting interviews in local languages in non-English speaking countries (LoCurto and Berg, 2016), and potential use of this scale for predicting participation in any future (COVID-19) vaccination programmes (Palamenghi et al., 2020).

What does this study add?

This is the first study to use Mokken scaling and IRT GRM in exploring properties of MM. Across countries, MM was shown to be unidimensional, with higher discrimination and broader cover of the latent trait differentially demonstrated among individual items. It seems that MM for doctors is more discriminating factor than for systemic aspects on organ donation. The two general items (1 and 3) form a unidimensional scale with the two-organ donation specific items (2 and 4). Therefore, even though the latter items are specific to organ donation, the scale could be used in different contexts

Limitations and Strengths

This study was conducted in the context of organ donation (Doherty et al., 2017; Doyle et al., 2019; O'Carroll et al., 2016) using the IRT parameters, so that it should be generalisable beyond

the sample in question (DeMars, 2011). Studies used brief questionnaire rather than interviews, a

combination might provide more information as has been suggested elsewhere (Feveile et al.,

2007). The inclusion criteria for the original studies were broad and the participants were

sampled from the general population, increasing generalisability. While we are limited by a lack

of data on ethnicity and small sample sizes from India and Malaysia, which limit sub-analyses

(Şahin and Anıl, 2017), the overall sample size is adequate for the approaches taken and results

should be generalisable beyond the samples obtained (Gregoire and Hambleton, 2009; Watson et

al., 2018). In the included studies, it was seen that items regarding doctors are more highly

discriminating in contrast to other studies and this needs to be understood in depth. Future

research should concentrate on comprehensive multidimensional MM scales with bigger

samples, and discrimination of MM between personalized and systemic aspects of healthcare

system, for example the recent anti-vaccination movement (Hussain et al., 2018).

CONCLUSIONS

Four MM items formed a highly discriminating unidimensional scale across countries and

measurement methods. An IRT GRM model showed that the measure captured moderate-high

levels of MM better than lower levels of MM, and that there was significantly higher item

discrimination for items concerning doctors, rather than systemic items. This MM measure

should have excellent generalisability across settings.

1   **REFERENCES**

2   Adams LB, Richmond J, Corbie-Smith G, et al. (2017) Medical Mistrust and Colorectal Cancer

3   Screening Among African Americans. *Journal of Community Health*. Springer New York

4   LLC. DOI: 10.1007/s10900-017-0339-2.

5   Ark LA van der (2007) Mokken Scale Analysis in R. *Journal of Statistical Software* 20(11).

6   Foundation for Open Access Statistic: 1–19. DOI: 10.18637/jss.v020.i11.

7   Benkert R, Cuevas A, Thompson HS, et al. (2019) Ubiquitous Yet Unclear: A Systematic

8   Review of Medical Mistrust. *Behavioral Medicine*. Routledge. DOI:

9   10.1080/08964289.2019.1588220.

10   Boone WJ, Yale MS and Staver JR (2014) *Rasch Analysis in the Human Sciences*. DOI:

11   10.1007/978-94-007-6857-4.

12   Chen WH, Revicki DA, Lai JS, et al. (2009) Linking Pain Items from Two Studies Onto a

13   Common Scale Using Item Response Theory. *Journal of Pain and Symptom Management*

14   38(4). NIH Public Access: 615–628. DOI: 10.1016/j.jpainsymman.2008.11.016.

15   Crutzen R and Peters GJY (2017) Scale quality: alpha is an inadequate estimate and factor-

16   analytic evidence is needed first of all. *Health Psychology Review* 11(3). Routledge: 242–

17   247. DOI: 10.1080/17437199.2015.1124240.

18   DeMars CE (2011) A Review of "The Theory and Practice of Item Response Theory".

19   *International Journal of Testing*. DOI: 10.1080/15305058.2011.556771.

20   Detailed Description of Data | organdonor.gov (n.d.). Available at:

21   https://www.organdonor.gov/learn/organ-donation-statistics/detailed-description#fig1

(accessed 14 November 2021).

DeVellis RF (2006) Classical test theory. *Medical Care* 44(11 SUPPL. 3). Med Care. DOI:
10.1097/01.mlr.0000245426.10853.30.

Dima AL (2018) Scale validation in applied health research: tutorial for a 6-step R-based
psychometrics protocol. *Health Psychology and Behavioral Medicine* 6(1). Informa UK
Limited: 136–161. DOI: 10.1080/21642850.2018.1472602.

Doherty S, Dolan E, Flynn J, et al. (2017) Circumventing the 'Ick' factor: A randomized trial of
the effects of omitting affective attitudes questions to increase intention to become an organ
donor. *Frontiers in Psychology* 8(AUG). Frontiers Media S.A. DOI:
10.3389/fpsyg.2017.01443.

Doyle F, Morgan K, Mathew M, et al. (2019) Theory content, question-behavior effects, or form
of delivery effects for intention to become an organ donor? Two randomized trials.
International Journal of Environmental Research and Public Health 16(7). MDPI AG.
DOI: 10.3390/ijerph16071304.

Edelen MO and Reeve BB (2007) Applying item response theory (IRT) modeling to
questionnaire development, evaluation, and refinement. In: *Quality of Life Research*,
August 2007, pp. 5–18. Springer. DOI: 10.1007/s11136-007-9198-0.

Fayers P (2004) Item Response Theory for Psychologists. *Quality of Life Research*. DOI:
10.1023/b:qure.0000021503.45367.f2.

Feveile H, Olsen O and Hogh A (2007) A randomized trial of mailed questionnaires versus
telephone interviews: Response patterns in a survey. *BMC Medical Research Methodology*

7. BioMed Central: 27. DOI: 10.1186/1471-2288-7-27.

Galvan FH, Bogart LM, Klein DJ, et al. (2017) Medical mistrust as a key mediator in the
association between perceived discrimination and adherence to antiretroviral therapy among
HIV-positive Latino men. *Journal of Behavioral Medicine* 40(5). Springer New York LLC:
784–793. DOI: 10.1007/s10865-017-9843-1.

Gaskin CJ and Happell B (2014) On exploratory factor analysis: A review of recent evidence, an
assessment of current practice, and recommendations for future use. *International Journal
of Nursing Studies*. Int J Nurs Stud. DOI: 10.1016/j.ijnurstu.2013.10.005.

Goold SD (2002) Trust, distrust and trustworthiness: Lessons from the field. *Journal of General
Internal Medicine*. Springer. DOI: 10.1046/j.1525-1497.2002.11132.x.

Gregoire J and Hambleton RK (2009) Advances in Test Adaptation Research: A Special Issue.
International Journal of Testing. DOI: 10.1080/15305050902880678.

Guadagnolo BA, Cina K, Helbig P, et al. (2009) Medical mistrust and less satisfaction with
health care among native americans presenting for cancer treatment. *Journal of Health Care
for the Poor and Underserved* 20(1). NIH Public Access: 210–226. DOI:
10.1353/hpu.0.0108.

Hambleton RK (2009) Applications of item response theory to improve health outcomes
assessment: developing item banks, linking instruments, and computer-adaptive testing. In:
Outcomes Assessment in Cancer. DOI: 10.1017/cbo9780511545856.022.

Hardin HK, McCarthy VL, Speck BJ, et al. (2018) Diminished Trust of Healthcare Providers,
Risky Lifestyle Behaviors, and Low Use of Health Services: A Descriptive Study of Rural

Adolescents. *Journal of School Nursing* 34(6). SAGE Publications Inc.: 458–467. DOI: 10.1177/1059840517725787.

Hardouin J-B (2004) MSP: Stata module to perform the Mokken Scale Procedure. *The Stata Journal* (June 2004). Boston College Department of Economics. Available at: https://ideas.repec.org/c/boc/bocode/s439402.html (accessed 30 July 2020).

Hays RD, Morales LS and Reise SP (2000) Item Response Theory and health outcomes measurement in the 21st century. *Medical Care* 38(9 SUPPL. 2). Lippincott Williams and Wilkins. DOI: 10.1097/00005650-200009002-00007.

Henrich J, Heine SJ and Norenzayan A (2012) The Weirdest People in the World? *SSRN Electronic Journal*. Elsevier BV. DOI: 10.2139/ssrn.1601785.

Hussain A, Ali S, Ahmed M, et al. (2018) The Anti-vaccination Movement: A Regression in Modern Medicine. *Cureus* 10(7). Cureus, Inc. DOI: 10.7759/cureus.2919.

Jabrayilov R, Emons WHM and Sijtsma K (2016) Comparison of Classical Test Theory and Item Response Theory in Individual Change Assessment. *Applied Psychological Measurement* 40(8). SAGE Publications Inc.: 559–572. DOI: 10.1177/0146621616664046.

Kolva E, Rosenfeld B, Liu Y, et al. (2017) Using item response theory (IRT) to reduce patient burden when assessing desire for hastened death. *Psychological Assessment* 29(3). American Psychological Association Inc.: 349–353. DOI: 10.1037/pas0000343.

Larson HJ, Clarke RM, Jarrett C, et al. (2018) Measuring trust in vaccination: A systematic review. *Human Vaccines and Immunotherapeutics*. Taylor and Francis Inc. DOI: 10.1080/21645515.2018.1459252.

LoCurto J and Berg GM (2016) Trust in healthcare settings: Scale development, methods, and preliminary determinants. *SAGE Open Medicine*. DOI: 10.1177/2050312116664224.

Miller J, Currie S and O'Carroll RE (2019) 'What if I'm not dead?' – Myth-busting and organ donation. *British Journal of Health Psychology* 24(1). John Wiley and Sons Ltd.: 141–158. DOI: 10.1111/bjhp.12344.

Mithra P, Ravindra P, Unnikrishnan B, et al. (2013) Perceptions and attitudes towards organ donation among people seeking healthcare in tertiary care centers of coastal south India. *Indian Journal of Palliative Care* 19(2). Wolters Kluwer -- Medknow Publications: 83–87. DOI: 10.4103/0973-1075.116701.

Montague E (2010) Validation of a trust in medical technology instrument. *Applied Ergonomics* 41(6). Elsevier Ltd: 812–821. DOI: 10.1016/j.apergo.2010.01.009.

Morgan SE, Stephenson MT, Harrison TR, et al. (2008) Facts versus 'feelings': How rational is the decision to become an organ donor? *Journal of Health Psychology* 13(5). SAGE Publications Ltd: 644–658. DOI: 10.1177/1359105308090936.

Morgan SE, Harrison TR, Afifi WA, et al. (2008) In their own words: The reasons why people will (not) sign an organ donor card. *Health Communication* 23(1). Health Commun: 23–33. DOI: 10.1080/10410230701805158.

National Organ & Tissue Transplant Organization (n.d.) Organ Report. Available at: https://notto.gov.in/organreport.htm (accessed 21 September 2020).

O'Carroll RE, Shepherd L, Hayes PC, et al. (2016) Anticipated regret and organ donor registration: A randomized controlled trial. *Health Psychology* 35(11). American

Psychological Association Inc.: 1169–1177. DOI: 10.1037/hea0000363.

Ockey GJ (2013) Item response theory. In: *The Routledge Handbook of Language Testing*. DOI: 10.4324/9780203181287-36.

Opel DJ, Lo B and Peek ME (2021) Addressing Mistrust About COVID-19 Vaccines Among Patients of Color. *Annals of Internal Medicine*. American College of Physicians. DOI: 10.7326/m21-0055.

Ozawa S and Sripad P (2013) How do you measure trust in the health system? A systematic review of the literature. *Social Science and Medicine*. Pergamon. DOI: 10.1016/j.socscimed.2013.05.005.

Palamenghi L, Barello S, Boccia S, et al. (2020) Mistrust in biomedical research and vaccine hesitancy: the forefront challenge in the battle against COVID-19 in Italy. *European Journal of Epidemiology* 1. Springer: 1. DOI: 10.1007/s10654-020-00675-8.

Pallant JF and Tennant A (2007) An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS). *British Journal of Clinical Psychology*. DOI: 10.1348/014466506X96931.

Pellowski JA, Price DM, Allen AM, et al. (2017) The differences between medical trust and mistrust and their respective influences on medication beliefs and ART adherence among African-Americans living with HIV. *Psychology and Health* 32(9). Routledge: 1127–1139. DOI: 10.1080/08870446.2017.1324969.

Petrillo J, Cano SJ, McLeod LD, et al. (2015) Using classical test theory, item response theory, and rasch measurement theory to evaluate patient-reported outcome measures: A

comparison of worked examples. *Value in Health* 18(1). Elsevier Ltd: 25–34. DOI: 10.1016/j.jval.2014.10.005.

Quinn KG, Kelly JA, DiFranceisco WJ, et al. (2018) The Health and Sociocultural Correlates of AIDS Genocidal Beliefs and Medical Mistrust Among African American MSM. *AIDS and Behavior* 22(6). Springer New York LLC: 1814–1825. DOI: 10.1007/s10461-016-1657-6.

Reckase MD (2009) *Multidimensional Item Response Theory*. DOI: 10.1007/978-0-387-89976-3.

Rose A, Peters N, Shea JA, et al. (2004) Development and Testing of the Health Care System Distrust Scale. *Journal of General Internal Medicine*. J Gen Intern Med. DOI: 10.1111/j.1525-1497.2004.21146.x.

Rothman AJ (2004) Guttman Scale. In: *Encyclopedia of Psychology, Vol. 4*. DOI: 10.1037/10519-019.

Russell E, Robinson DHZ, Thompson NJ, et al. (2012) Distrust in the healthcare system and organ donation intentions among African Americans. *Journal of Community Health* 37(1). NIH Public Access: 40–47. DOI: 10.1007/s10900-011-9413-3.

Sahay M (2019) Men are from mars, women are from venus: Gender disparity in transplantation. *Indian Journal of Transplantation* 13(4). Medknow: 237. DOI: 10.4103/ijot.ijot_72_19.

Şahin A and Anıl D (2017) The effects of test length and sample size on item parameters in item response theory. *Kuram ve Uygulamada Egitim Bilimleri*. DOI: 10.12738/estp.2017.1.0270.

Samejima F (2016) Graded response models. In: *Handbook of Item Response Theory*. DOI: 10.1201/9781315374512.

Saunders J (2017) Trust and mistrust between patients and doctors. In: *Handbook of the*

Philosophy of Medicine. Springer Netherlands, pp. 487–502. DOI: 10.1007/978-94-017-8688-1_20.

Sax LJ, Gilmartin SK and Bryant AN (2003) Assessing response rates and nonresponse bias in web and paper surveys. *Research in Higher Education*. DOI: 10.1023/A:1024232915870.

Shelton RC, Winkel G, Davis SN, et al. (2010) Validation of the group-based medical mistrust scale among urban black men. *Journal of General Internal Medicine* 25(6). Springer New York: 549–555. DOI: 10.1007/s11606-010-1288-y.

Sheppard VB, Huei-Yu Wang J, Hurtado-de-Mendoza A, et al. (2019) Psychometric Properties of the Medical Mistrust Index (MMI) in Latina Immigrants. *Behavioral Medicine* 45(2). Routledge: 128–133. DOI: 10.1080/08964289.2019.1585326.

Sijtsma K and Molenaar I (2011) *Introduction to Nonparametric Item Response Theory*. SAGE Publications, Inc. DOI: 10.4135/9781412984676.

Smirnoff M, Wilets I, Ragin DF, et al. (2018) A paradigm for understanding trust and mistrust in medical research: The Community VOICES study. *AJOB Empirical Bioethics* 9(1). Taylor and Francis Inc.: 39–47. DOI: 10.1080/23294515.2018.1432718.

StataCorp (2017) *STATA Item Response Theory Reference Manual, Release 15*.

StataCorp (2019) Stata: Software for Statistics and Data Science. *Stata Statistical Software*. Available at: https://www.stata.com/ (accessed 30 July 2020).

Sullivan LS (2020) Trust, Risk, and Race in American Medicine. *Hastings Center Report* 50(1). John Wiley and Sons Inc.: 18–26. DOI: 10.1002/hast.1080.

Swaminathan H, Hambleton RK and Rogers HJ (2006) 21 Assessing the Fit of Item Response

Theory Models. *Handbook of Statistics*. Elsevier. DOI: 10.1016/S0169-7161(06)26021-8.

Taber JM, Leyva B and Persoskie A (2015) Why do People Avoid Medical Care? A Qualitative Study Using National Data. *Journal of General Internal Medicine* 30(3). Springer: 290–297. DOI: 10.1007/s11606-014-3089-1.

Tarka P (2018) An overview of structural equation modeling: its beginnings, historical development, usefulness and controversies in the social sciences. *Quality and Quantity* 52(1). Springer Netherlands: 313–354. DOI: 10.1007/s11135-017-0469-8.

The R Foundation (2018) R: The R Project for Statistical Computing. Available at: https://www.r-project.org/ (accessed 30 July 2020).

Thompson HS, Valdimarsdottir HB, Winkel G, et al. (2004) The Group-Based Medical Mistrust Scale: Psychometric properties and association with breast cancer screening. *Preventive Medicine* 38(2). Academic Press Inc.: 209–218. DOI: 10.1016/j.ypmed.2003.09.041.

van Schuur WH (2003) Mokken Scale Analysis: Between the Guttman Scale and Parametric Item Response Theory. *Political Analysis*. DOI: 10.1093/pan/mpg002.

Vincent B, Kumar G, Parameswaran S, et al. (2019) Barriers and suggestions towards deceased organ donation in a government tertiary care teaching hospital: Qualitative study using socio-ecological model framework. *Indian Journal of Transplantation* 13(3). Medknow: 194. DOI: 10.4103/ijot.ijot_9_19.

Watson R, Van der Ark LA, Lin LC, et al. (2012) Item response theory: How Mokken scaling can be used in clinical practice. *Journal of Clinical Nursing* 21(19–20). J Clin Nurs: 2736–2746. DOI: 10.1111/j.1365-2702.2011.03893.x.

1   Watson R, Egberink IJL, Kirke L, et al. (2018) What are the minimal sample size requirements

2   for mokken scaling? An empirical example with the warwick- edinburgh mental well-being

3   scale. *Health Psychology and Behavioral Medicine*. DOI:

4   10.1080/21642850.2018.1505520.

5   Williams DR and Mohammed SA (2009) Discrimination and racial disparities in health:

6   Evidence and needed research. *Journal of Behavioral Medicine* 32(1). J Behav Med: 20–47.

7   DOI: 10.1007/s10865-008-9185-0.

8   Williamson LD and Bigman CA (2018) A systematic review of medical mistrust measures.

9   *Patient Education and Counseling*. Elsevier Ireland Ltd. DOI: 10.1016/j.pec.2018.05.007.

10   Xia J, Tang Z, Wu P, et al. (2019) Use of item response theory to develop a shortened version of

11   the EORTC QLQ-BR23 scales. *Scientific Reports* 9(1). Nature Publishing Group: 1–10.

12   DOI: 10.1038/s41598-018-37965-x.

13   Yang FM and Kao ST (2014) Item response theory for measurement validity. *Shanghai Archives*

14   *of Psychiatry* 26(3). Shanghai Mental Health Center: 171–177. DOI: 10.3969/j.issn.1002-

15   0829.2014.03.

16

17

1 **FIGURES AND TABLES**

2 **Table 1: Mokken scaling results for total sample**

Item	N	Mean score	Observed Guttman errors	Expected Guttman Errors	Loevinger H-coefficient	z-statistic	p	Number of non-significant H
mm1	1460	2.79	6889	12929.44	0.46719	29.2225	<.001	0
mm3	1460	2.95	6302	13297.85	0.52609	32.9586	<.001	0
mm2	1460	2.28	6037	12704.04	0.52480	32.5848	<.001	0
mm4	1460	2.44	6102	13281.47	0.54056	34.0639	<.001	0
Scale	1460		12665	26106.41	0.51487	45.5618	<.001	-

3 **Table 2: Item discrimination scores**

| Items	Discrimination	95% CI
MM1 = 'Hospitals sometimes prescribe medication as a way of experimenting on people without their knowledge of consent'	1.49	1.32-1.67
MM2 = 'If I sign an organ donor card, doctors might take my organs before I'm actually dead'	2.78	2.44-3.11
MM3 = 'Sometimes, medical procedures are done on people without their consent'	1.94	1.73-2.15
MM4 = 'If I sign an organ donor card, doctors might not try so hard to save my life'	3.12	2.71-3.53

1

Figure 1: Test level information
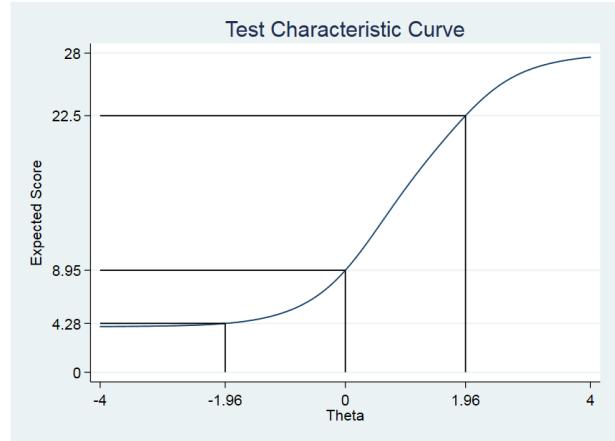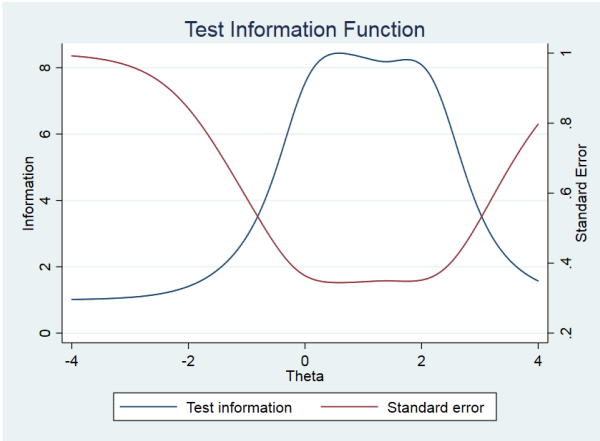
1.1 Test information function   **Figure 1.2: Test characteristic curve**
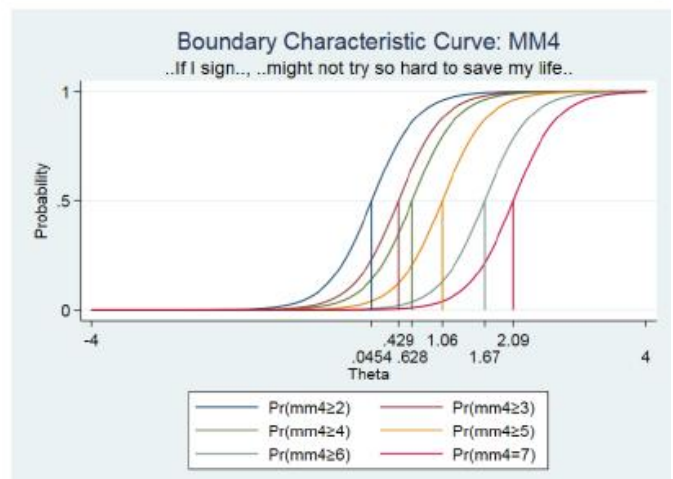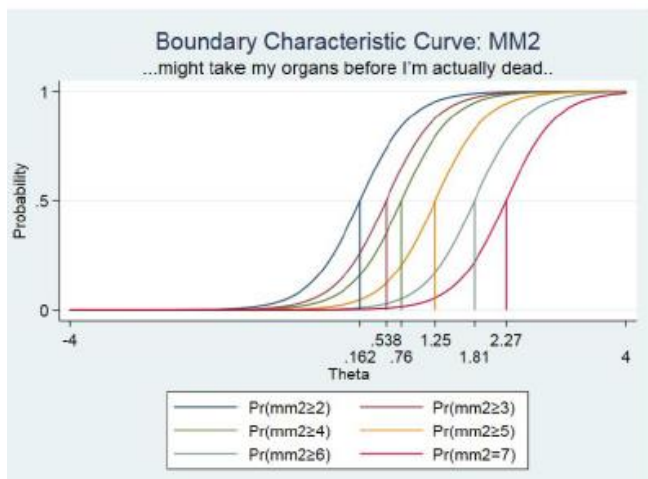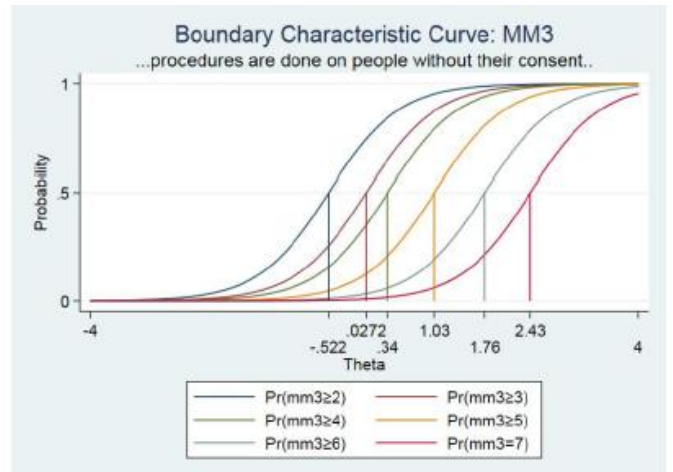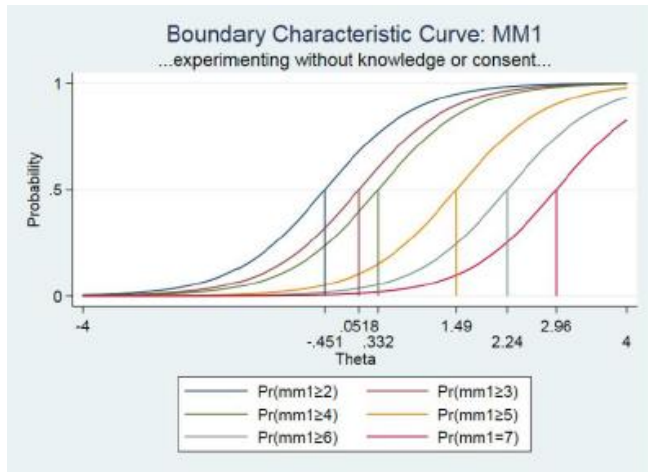
Figure 2: Item level information

4
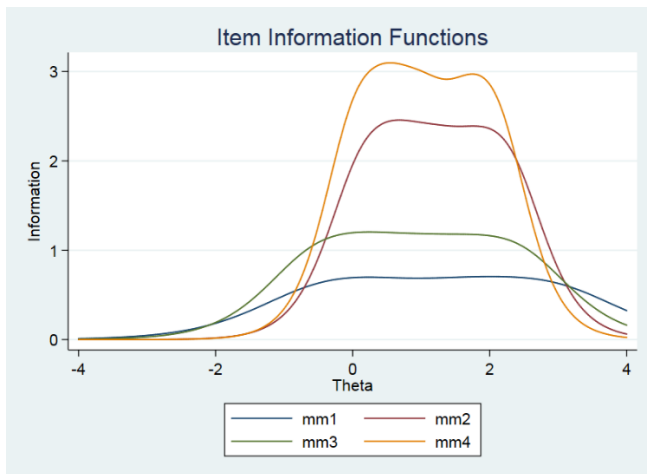
2.1 Item Information Curves

6

Figure 2.2: Item Characteristic Curve for each of the 4 items
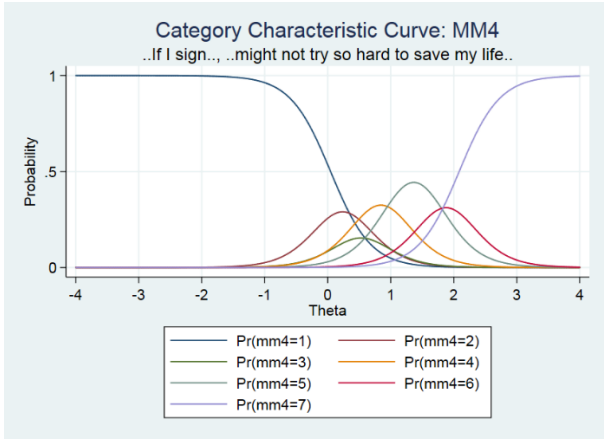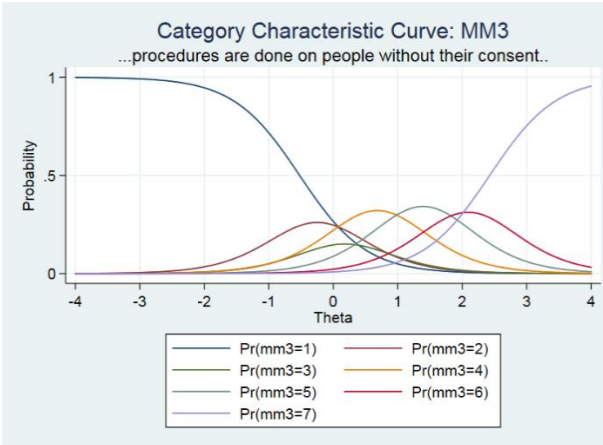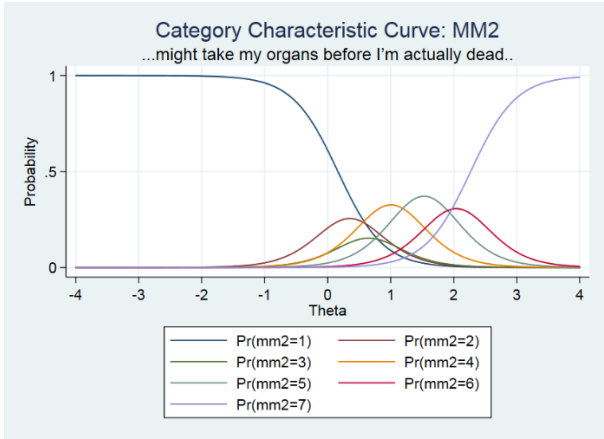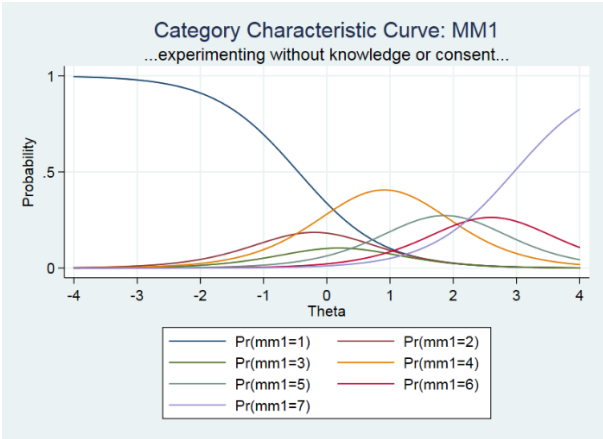
Figure 3. Category characteristic curves (CCCs) for each of the 4 items

1 **APPENDIX**

2 1. The model fit statistics (Akaike information criterion (AIC) /Bayesian information criteria

3 (BIC)) for IRT models were as follows (the lowest scores indicate the best fit):

4 Nominal response model   16945.98, 17200.21

5 Partial credit model   17095.25,17227.66

6 Rating scale model   17194.33,17247.29

7 Graded response model   16857.49,17005.79

8 2. The graded response model can be summarized by the equation

9 *P(Xij = x ij | θ i) = P*xij(θi) – P*xij+1 (θi)*

$$P_{x_{ij}}^*\left(\theta_i\right) = P\left(X_{ij} \geq x_{ij} \mid \theta_i\right) = \frac{e^{Da_j\left(\theta_i - b_{x_{ij}}\right)}}{1 + e^{Da_j\left(\theta_i - b_{x_{ij}}\right)}}$$

10 Where:

11 • θ represents the latent ability or trait, and its actual level in the test subject.

12 • xij represents the grade given.

13 • bjx is a constant specific to the test item; the location parameter, or category boundary for

14   score x; the point on the ability scale where p = 0.5

15 • ajx is another constant specific to the test item, the discrimination parameter, and is

16   constant over response categories for a given item.

17 • D is a scale factor

18

1   2. Correlation matrix of medical mistrust items

	mm1	mm2	mm3	mm4
mm1	1.0000000			
mm2	0.3698131	1.0000000		
mm3	0.5303012	0.4728881	1.0000000	
mm4	0.4179939	0.6377098	0.4982345	1.0000000

2