Received: 15 June 2016 | Revised: 2 October 2016 | Accepted: 20 October 2016

DOI: 10.1002/mpr.1551

WILEY

ORIGINAL ARTICLE

Evaluating psychological distress data

James McIntosh[1,2]

[1] Economics Department, Concordia University, Montreal, Canada

[2] 623 Econometrics, Montreal, Canada

Correspondence
Economics Department, Concordia University, 1455 De Maisonneuve Blvd. W., Montreal, H3G 1M8, Quebec, Canada.
Email: james.mcintosh@concordia.ca

Abstract

Kessler k_6 psychological distress scores are analyzed using a count model and item response theory (IRT) models are applied to the items which produce the k_6 score and generate an alternative distress score, θ^*. Other ways of utilizing the constituent items are also examined. The data used in the analysis comes from the 2014 National Survey of Drug Use and Health. Three important results emerge. First, θ^* and k_6 are not highly correlated and their distributions are quite different. The k_6 score gives a much more favourable picture of mental health than θ^*. Second, k_6 does a much better job in explaining participation in treatment programs than θ^* suggesting a very limited role for IRT methods in the analysis of psychological distress data. As a diagnostic tool k_6 is an effective and simple way of summarizing the item data. Third, for researchers interested in which individual characteristics determine psychological distress better results are obtained by analyzing the six constituent items which are used to generate the k_6 score using ordered probability models rather than k_6 itself.

KEYWORDS

Item response models, National Survey of Drug Use and Health, Psychological Distress Score

1 | INTRODUCTION

The Kessler k_6 psychological distress score, Kessler et al. (2001), Kessler et al. (2002), and Kessler et al. (2010) and the World Health Organization Disability Assessment Schedule, WHODAS 2.0 (World Health Organization [WHO], 2010), are non-specific screening scales for psychological distress. They are used to identify patients with serious mental illness and are operational in many countries. They are based on six or eight questions that ask respondents to evaluate how they feel about themselves in terms of how much anxiety or lack of self-worth they are experiencing. The list of specific questions for the Kessler score appears in Section 3. These questions are summarized by the k_6 or the WHODAS score which is just the sum of the item category answers for each respondent. Answers to questions are frequencies of having a particular attribute, a feeling of hopelessness, for example. Respondents can give five possible responses: none of the time, a little of the time, to all of the time. In the National Survey of Drug Use and Health (NSDUH) the category codes are 1 for all of the time and 5 for none of the time. Usually codes run the other way from 0 to 4 and this gives the usual k_6 score which runs from 0 to 24 with high scores indicating serious psychological distress.

Because of the way the item outcomes are coded and the additive nature of the k_6 score it has some undesirable properties. First, using category codes running from 1 to 5 in a regression model imposes the condition that differences in the severity of the distress between categories are the same no matter what category is being considered. This is not a very plausible condition to impose on the item data and it is not supported by statistical models that are designed to analyze this type of data (see Greene, 2008, chapter 23). Ordered probability models with a set of equally spaced threshold points[1] are always rejected in favor of models with unequally spaced threshold points when these are applied to the individual items. As one would expect in terms of the importance of an increase in a psychological attribute, going from none of the time to a little of the time is not the same as going from most of the time to all of the time.

Secondly, there many different outcomes that end up having the same k_6 score. This leads to a lack of precision in evaluating individual psychological distress. Patients with low levels of distress on many items can get the same score as a patient with high levels of distress on a few of the items.

Thirdly, there is also a lack of precision that arises because the items are highly correlated. This effectively involves some double counting and can lead to inflated scores.

Fourth, the procedure gives equal weights to the six items. This assumes that they are of equal importance in determining the type

[1] Threshold points usually refer to severity levels and increase the more time spent with the malady. They are estimated as parameters in item and IRT models.

of treatment a respondent receives, a result which is not supported by the data.

Some of these issues have been recognized by researchers in the area and this has led to the use of item response theory (IRT) models applied to the individual items which serve as the basis for these aggregate scores (see for, example, Kessler et al. [2010], and Kryner, Osborne, Duck, Houkamou, and Sibley [2013]). While this procedure could be seen as an improvement over using crude aggregate scores since the first three problems noted earlier disappear when an IRT model is used, it has not replaced the Kessler and WHO scores in the process of screening for serious mental illness. IRT results have mainly been used to improve the k_6 score by using the severity (threshold) parameters to weight the higher level outcomes. But most researchers find that unweighted and weighted scores are highly correlated.

In addition there is a score which can be derived directly from an IRT model which is the Bayesian conditional mean, θ^* first proposed by Lindsay, Clogg, and Crego (1991). This an alternative to k_6 and can be compared to k_6 in terms of its ability to explain which respondents get treatments as well as which regressors are important in explaining psychological distress.

The fourth problem turns out to be quite serious and requires alternative procedures to address it. In addition to the psychological scores the NHDUS also asks respondents whether they received any treatment for mental illness. This data is analyzed using the two scores already mentioned but two alternative representations of the psychological item are also considered. First k_6 is disaggregated into its constituent sub-scores $\{i_j\ j = 1, 2 \ldots 6\}$ where, of course,

$k_6 = \sum_{j=1}^{6} i_j.$

Using the sub-scores tests the hypothesis that all of the items are equally important. As will be seen there is considerable variation in the importance of sub-scores and use of k_6 suppresses these differences. The second representation of the item data is the vector $D = \{d_h, h = 1, 2 \ldots 24\}$ of dummy variables for the individual non-zero item outcomes. This is the most general way of representing the item data in the analysis of treatments.

The paper is organized in the following way. First an IRT model is applied to the six items used to generate the k_6 scores. The technical details concerning the model are outlined in the Appendix. Its relative performance is evaluated by comparing its statistical characteristics with those of k_6. Other alternatives to k_6 are then explored. The results of this exercise are shown later where all measures of psychological distress are evaluated in terms of their ability to explain participation in treatment programs. The next section describes the data used in the analysis.

2 | DATA

The data comes from the 2014 NSDUH. The k_6 score is based on a set of items which represent the psychological problems that afflict the respondents. They are asked how often they experienced over the last month the following six conditions. The answers run from none of the time to all of the time and there are five gradations in the possible replies.

The items are:

1. How often did you feel nervous?
2. How often did you feel hopeless?
3. How often did you feel restless or fidgety?
4. How often did you feel so depressed that nothing could cheer you up?
5. How often did you feel that everything was an effort?
6. How often did you feel worthless?

They are scored from 0 to 4 where 4 represents all of the time. The k_6 score is just the sum of these six item scores so it ranges from 0 to 24. The information in Table 1 gives k_6 scores by gender and age-group.

As well as information on the respondent's psychological state the survey elicits demographic data. This includes age, marital status, income, educational attainment as well as data on smoking and alcohol use. The specific respondent variables are listed in Table 2. There are also three treatment variables which are denoted as T_1, T_2, and T_3. The first is a dummy variable which takes the value one if the respondent had any medicine prescribed for mental health issues. The second indicates mental health outpatient treatment, and the third is for any mental health treatment. The survey is quite large so that models can be applied to both age-group and gender specific data.

As the data in Table 1 shows, there is an obvious need to disaggregate the data by gender and age group as k_6 scores differ across these two classifications. The most distinguishing features of the data in Table 1 is the decline in psychological distress as the population ages and the lower rates of distress for men.

TABLE 1 Average k_6 scores for males and females by age group

	Males			Females		
Age group	Mean	Sample size	Percentage	Mean	Sample size	Percentage
	k_6		$k_6 = 0$	k_6		$k_6 = 0$
18–23	4.94	6132	0.17	5.92	6686	0.12
24–34	4.01	3763	0.23	4.63	4512	0.18
35–49	3.50	5208	0.28	4.23	5891	0.21
50–64	3.06	2415	0.33	3.78	2902	0.26
64+	2.51	1623	0.38	3.03	1943	0.29
Average/sum	3.92	19141	0.25	4.66	21941	0.19

TABLE 2 IRT and negative binomial regression parameter estimates: males aged 24–34

	IRT model	Negative binomial model
ln(income)	−0.19 (0.01)	−0.16 (0.02)
Race		
Black	0.08 (0.04)	0.10 (0.06)
Asian	0.09 (0.06)	0.08 (0.08)
Hispanic	0.06 (0.03)	−0.10 (0.05)
Other	0.15 (0.06)	0.17 (0.08)
Education		
High school graduate	0.04 (0.04)	0.09 (0.06)
Some college	0.19 (0.04)	0.24 (0.06)
College graduate	0.26 (0.04)	0.32 (0.07)
Smoking behavior (cigarettes per day)		
<1	0.19 (0.03)	−0.04 (0.11)
2–5	0.02 (0.08)	0.21 (0.09)
6–16	0.14 (0.04)	0.14 (0.06)
16–25	0.17 (0.04)	0.14 (0.06)
26–35	0.25 (0.07)	0.17 (0.07)
36+	0.87 (0.16)	0.50 (0.16)
Alcohol use (days per month)		
1–7	−0.08 (0.03)	−0.09 (0.04)
8–12	−0.01 (0.04)	−0.12 (0.06)
13–18	−0.03 (0.06)	−0.17 (0.08)
19–23	0.04 (0.05)	0.04 (0.08)
24–30	0.18 (0.06)	−0.02 (0.09)
Dispersion parameter	σ 1.90(0.03)	b 1.05 (0.02)

3 | RESULTS

The IRT item severity and variance parameters are shown in Table 3 for the male age group 24–34.

Other age groups were examined but produce qualitatively similar results, although the parameter estimates are not the same for all gender specific age groups. The Verhelst–Glas (Verhelst & Glas, 1995) normalization rule is used to identify the parameters.

Under this normalization rule one of the severity and one of the variance parameters has to be chosen arbitrarily; δ_1 and σ_1 are set equal to 0 and 1, respectively. As required, the severity increase with the category. These results are similar to those found using Australian data by Kryner, Osborne, Duck, Houkamou, and Sibley (2013). With the exception of the felt worthless item the variance terms are all significantly different from unity.

Table 2 displays the regression parameter estimates for both the item response model and the negative binomial count model for the variable k_6. This distribution is well suited for analyzing the data since it is over-dispersed (the mean of k_6 is substantially less than its variance). It fits the data much better than the linear regression model.

Two significant results emerge from Table 2. First, the estimated coeffcients are similar for both methods. Secondly, having a higher income, consuming moderate amounts of alcohol, not smoking all lead to better mental health. But contrary to what might be expected being better educated does not.

Where the two methods differ is in the score distributions which are displayed in Table 4 for each method by quintile for the age group 24–34. The score for the IRT model is the conditional mean of θ, $E(\theta|y)$. As Table 4 indicates the k_6 score gives a much lower average level of psychological distress than the IRT score. The two scores are quite different and the correlation between the two scores is 0.79 for this age group.

In Table 5 the ability of the all of methods to explain who received treatment for mental health problems is analyzed.

The mental health treatment variables are labeled T_1 to T_3. The criterion used to evaluate each model is McFadden's R^2. This measures the percentage increase in the ln-likelihood function over baseline which is a model which has no regressors other than an intercept term. In the first column only X, the set of variables which describe the respondent's characteristics and are those which appear in Table 2, are used to explain the treatments. The second column uses θ^* as the regressor. Furthermore, X are not included as regressors since they are already included in θ^*. The R^2 coeffcient for X and k_6 in the third column is significantly larger than that for θ^*. Thus, k_6 does a much better job in explaining who receives treatments than θ^* in spite of the earlier noted limitations in the way it is constructed. The reasons for this will be explained later. In column 4 the components of k_6, i_1 to i_6, and X are used as regressors. These do significantly better than k_6. It is also the case that the regression coeffcients differ substantially over the sub-scores, a point which will be helpful in the later discussion of how θ^* performs. The individual item dummies are the best performers and models involving them should be used if the objective is to determine which respondent characteristics are most important in determining who receives

TABLE 3 Estimated category severity and variance parameters: males aged 24–34

Psychological distress item	j	κ_{j1}	κ_{j2}	κ_{j3}	κ_{j4}	σ_j
Felt nervous	1	0.0	−0.38 (0.06)	0.63 (0.06)	1.74 (0.06)	1.0
Felt hopeless	2	−0.38 (0.07)	−0.10 (0.06)	0.90 (0.06)	2.02 (0.06)	0.58 (0.02)
Felt restless	3	−0.34 (0.07)	0.12 (0.06)	1.12 (0.06)	2.24 (0.06)	0.71 (0.03)
Felt depressed	4	−0.51 (0.08)	0.04 (0.06)	1.05 (0.06)	2.17 (0.07)	0.74 (0.02)
Everything was an effort	5	0.79 (0.08)	1.80 (0.07)	2.81 (0.07)	3.93 (0.07)	1.20 (0.02)
Felt worthless	6	−0.61 (0.09)	0.08 (0.06)	1.09 (0.06)	2.21 (0.09)	1.05 (0.04)

The categories are: 1 is all of the time, 2 is most of the time, 3 is some of the time, 4 is a little of the time and 5 is none of the time. Standard errors are in brackets.

TABLE 4 Score distributions for θ^* and k_6

Quintile	θ^*	k_6
Q1	0.45	0.60
Q2	0.29	0.22
Q3	0.15	0.11
Q4	0.11	0.05
Q5	0.01	0.02
Sum	1.0	1.0

Note: Quintile 1 (Q1) represents respondents with the lowest level of psychological distress for both indexes.

TABLE 5 Goodness-of-fit statistics for all treatment models: males aged 24–34

	R^2				
	X	θ^*	X,k_6	X,I	X,D
T_1	0.07	0.10	0.18	0.19	0.21
T_2	0.08	0.12	0.19	0.20	0.23
T_3	0.07	0.10	0.18	0.19	0.20

Note: R^2, McFadden's R-squared; X is a vector of respondent characteristics; I is the vector of sub-scores; D is the vector of item dummies.

treatment. It should also be noted that the increases in R^2 are significant as one moves from left to right in Table 5.

Table 6 shows that the effects of the individual sub-scores are not the same in explaining treatments and what is particularly interesting is that not all of the six items are significant in explaining them. Having the second malady actually reduces the probability of all three treatments.

The validity of k_6 as a reliable measure of mental health depends on what use is being made of it. Researchers who are interested in which respondent characteristics are important in determining phycological distress should apply ordered probability models to the individual items. This is more informative than trying to explain k_6 and, as already noted, what matters depends on which item is being examined. But as a diagnostic tool, k_6 is an effective and simple way of summarizing the item data and it outperforms θ^*. It looses some of its precision due the inclusion of non-significant items. Because of this practitioners might like to examine some of the individual item outcomes in addition to k_6. On balance, however, there is no reason not to continue using k_6 as a diagnostic tool.

TABLE 6 Probability model parameter estimates for the effects of i_1 to i_6 on mental health treatment indicators, ages 24–34

	T_1	T_2	T_3
i_1	−0.46(0.08)	−0.30(0.10)	−0.41(0.08)
i_2	0.023 (0.11)	0.06 (0.13)	0.11 (0.10)
i_3	−0.26(0.08)	−0.22(0.09)	−0.23(0.07)
i_4	−0.40(0.11)	−0.39(0.12)	−0.44(0.10)
i_5	−0.09 (0.07)	−0.19(0.09)	−0.09 (0.08)
i_6	−0.15 (0.10)	−0.14 (0119)	0.09 (0.09)

4 | DISCUSSION AND CONCLUSIONS

As Table 6 shows items 1, 3, and 4 are the items that matter when treatments are being considered. The others are not significant. The individual items also explain much more of the variation in respondent participation in treatment programs than θ^* suggesting a rather limited role for IRT methods in the analysis of psychological distress. There are no computational costs associated with its use and that makes it accessible to a wide range of health professionals who do not have access to or the competence to use sophisticated statistical software.

IRT models avoid some of the problems associated with additive representations of item information. The question then arises as to why they perform so poorly compared to k_6 or alternative measures of psychological distress. The first reason is that it imposes restrictions that are not supported by the data. The IRT model assumes that the regression parameters are the same for all items, an assumption which is not supported by the application of ordered probability models to the individual items. Although individual variance parameters can be identified and are allowed to differ across the items there are still differences in the item parameter estimates that are not accounted for by introducing item specific variance parameters.

There is a second more subtle reason. In the Appendix Equation A4 gives the components of θ^*; there is a mean function which depends on the variables describing respondent characteristics and the conditional mean of the random effect. Here the regressors are not particularly informative about who receives treatment, as column 1 of Table 5 shows. The random effect contributes more to the variation in treatments than the mean but it also is not a good substitute for the item information itself. Although k_6 is not the best way of using the item information to explain participation in treatments it does quite well and does not depend on regressors. This explains why X and k_6 explain so much more of variation in $\{T_k, k = 1, 2, 3\}$ than θ^*.

5 | APPENDIX: ITEM RESPONSE MODELS

Item response models have been used extensively in the analysis of educational test scores. Discussion of the method begins a brief review of how test scores are analyzed. Some of the more important recent references for this work are Bock and Moustaki (2007), von Davier and Carstensen (2007), Downing and Haladyna (2006), and Aitkin and Aitkin (2011).

To understand the content of this theory some notation is needed. Let there be V individuals and let the ability of individual v be θ_v. Consider a group or cluster of J test questions, or items as they are referred to in the literature. Let the vector $y_v = (y_{v1}, y_{v2} \ldots y_{vJ})$ where $y_{vj} \in \{0,1\}$ are the values of v responses to the J items. In this "dichotomous" case there is just one right answer and getting the correct answer for item j makes $y_{vj} = 1$. There is a long standing tradition in the test score literature of attributing item success to the difference between the individual's ability and the item's diffculty. In this framework item responses are generated by a latent variable, y^*_{vj}, crossing a diffculty threshold δ_j. This

latent variable is defined as

$$y^*_{vj}, = \theta_v + u_{vj} \tag{A1}$$

where

$$\theta_v = \mu_v(\beta, X_v) + \varepsilon_v \tag{A2}$$

and

$$\mu_v(\beta, X_v) = \beta_0 + \sum_{k=1}^{K} x_{vk}\beta_k \tag{A3}$$

when there are regressors present. It has a component which depends on the individuals personal characteristics as well as a random component which represents individual characteristics that are not observable to the researcher.

The Lindsay, Clogg, and Crego (1991) score associated with the IRT model, which is the conditional expectation of θ_v given the item outcomes is defined as

$$\theta^*_v = \mu_v(\beta, X_v) + E(\varepsilon_v|y_v) \tag{A4}$$

Equations A2 and A3 are the same equation as (4.14) in Adams and Wu (2007). This is an error components model where ε_v and u_{vj} are independent, $E(\varepsilon_v) = E(u_{vj}) = 0$, and the distribution of u_{vj} depends on the characteristics of item j. The u_{vj} values are independent and independent from Φ_v, and δ_j is the level of diffculty of item j. Individual v will answer item j correctly if $y^*_{vj} > \delta_j$. Conditional item probabilities are then defined as

$$\Pr\{y_{vj} = 0|\varepsilon_v\} = \Pr\{u_{vj} \leq \delta_j - \mu_v(\beta, X_v) - \varepsilon_v\}$$

$$= \Phi(\delta_j - \mu_v(\beta, X_v) - \varepsilon_v, \sigma_j)$$

$$= \Phi_{vj}(\varepsilon_v) \tag{A5}$$

where Φ is the cumulative normal distribution, σ^2_j is the variance of u_{vj}, X_v is a vector of the characteristics of v, and β is a vector of regression parameters which is the same for all individuals.[2] In this model the notions of ability and diffculty are separate. Ability does not depend on the item and item diffculty is the same for all individuals.

The conditional likelihood function for this model is

$$L(\beta, \delta, \sigma|\varepsilon) = \prod_{v=1}^{V}\prod_{j=1}^{J}\varphi_{vj}^{\left(1-yvj\left(1-\varphi_{vj}\right)\right)^{yvj}} \tag{A6}$$

Since the random effect in ability, ε_v, is not observable the average or integrated likelihood function is required for estimation purposes.

This is

$$L(\beta, \delta, \sigma) = \prod_{v=1}^{V}\int\prod_{j=1}^{J}\varphi_{vj}^{\left(1-yvj\left(1-\varphi_{vj}\right)\right)^{yvj}\phi(\varepsilon_v)d\varepsilon_v} \tag{A7}$$

where $\phi(\varepsilon_v)$ is the normal probability density function with mean zero and variance σ^2.

The asymmetry in the treatment of the two errors may appear strange to some readers who might have expected them to have been combined as $w_{vj} = \varepsilon_v + u_{vj}$. But then the w_{vj} errors are correlated across items and the likelihood function in Equation A5 would no longer be correct. The conditional independence assumption has major computational advantages and numerically integrating Equation A6 is much easier than trying to find a suitable multivariate distribution for w_{vj}.[3]

The IRT model used earlier to explain educational test scores needs to be altered to accommodate the multiple response nature of psychological items. In the Kessler model there are six psychological distress items and each has five ordered responses. The first item asks the amount of time over the last month the respondent felt nervous. Answers are none of the time, a little of the time, up to all of the time. Since these answers are ordered an ordered probability model can be used to analyze the response data. Again let the responses for item j be y_{vj} then

$$\Pr\{y_{vj} = 1|\varepsilon_v\} = \Phi(\kappa_{j1} - \mu_v(\beta, X_v) - \varepsilon_v)$$

$$\Pr\{y_{vj} = k|\varepsilon v\} = \Phi(\kappa_{jk} - \mu_v(\beta, X_v) - \varepsilon v) - \Phi(\kappa_{j(k-1)} - \mu_v(\beta, X_v) - \varepsilon_v)k = 2, 3, 4$$

$$\Pr\{y_{vj} = 5|\varepsilon v\} = 1 - \Phi(\kappa_{j4} - \mu_v(\beta, X_v) - \varepsilon v) \tag{A8}$$

There is a set of these five equations for each of the six items. The system looks much the same as the educational item response model except that the probabilities are cast in terms of distribution function differences and the threshold parameters are increasing for each item unlike the situation in the education model where the thresholds are the diffculty parameters which are ordered by proportion of correct answers for the item. The threshold parameters $\{\kappa_{je}\}$ are referred to as severity parameters. This model is similar to that used by Sibley (2012) or Kryner et al. (2013). Regression and severity parameter estimates are very similar but the maximized value of the likelihood function is significantly higher indicating that it fits the data better.

Possible differences in the importance of the items was mentioned in the Introduction. No account of this is taken in the procedures employed here. The items could be weighted in the likelihood function but it is not clear how this should be done.

ACKNOWLEDGMENTS

The author wishes to acknowledge helpful comments from Ronald Kessler and two anonymous reviewers.

[2]Much of the research involving item response models uses a logistic distribution instead on the normal distribution used here. Which distribution is actually used is not an issue in the literature.

[3]Numerical procedures for evaluating this integral may be found in Aitkin and Aitkin (2011), p. 34).

DECLARATION OF INTEREST STATEMENT

The author has no conflicts of interest.

REFERENCES

Adams, R. J., & Wu, M. L (2007). The mixed-coefficient multinomial logit model: A generalized form of the Rasch model. In M. von Davier & C. H. Carstensen, *Multivariate and Mixture Distribution Rasch Models: Extensions and Applications*. New York: Springer Science + Business Media.

Aitkin, M., & Aitkin, I. (2011). *Statistical modeling of the national assessment of educational progress. Statistics for Social and Behavioral Sciences*, 23.

Bock, D. D., & Moustaki, I. (2007). Item response theory in a general framework. In C. R. Rao, & S. Sinharay (Eds.), *Handbook of Statistics: Vol. 26. Psychometrics*), pp. 469–513. Amsterdam: North Holland Publishing.

Downing, S. M., & Haladyna, T. M. (2006). *Handbook of Test Development*. Mahwah NJ: Lawrence Erlbaum Associates.

Greene, W. H. (2008). *Econometric Analysis* (6th ed.). Pearson Prentice Hall, NJ, USA: Upper Saddle River.

Kessler, R. C., Berglund, P. A., Bruce, M. L., Koch, J. R., Laska, E. M., Leaf, P. J., … Wang, P. S. (2001). The prevalence and correlates of untreated serious mental illness. *Health Services Research*, *36*, 987–1007.

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S.-L. T., … Zaslatsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine*, *32*, 959–976.

Kessler, R. C., Green, J. G., Gruber, M. J., Sampson, N. A., Bromet, E., Cuitan, M., … Zaslavsky, A. M. (2010). Screening for serious mental illness in the general population with the K6 screening scale: Results from the WHO World Mental Health (WHM) Survey. *International Journal of Methods in Psychiatric Research*, *19*, 4–22.

Kryner, A. M., Osborne, D., Duck, I. M., Houkamou, C., & Sibley, C. M. (2013). Measuring psychological distress in New Zealand: Item response properties and demographic differences in the Kessler-6 screening measure. *New Zealand Journal of Psychology*, *42*, 69–83.

Lindsay, B. G., Clogg, C. L., & Crego, J. (1991). Semiparametric estimation in the Rasch Model and related exponential response models, including a simple latent class model for item analysis. *Journal of the American Statistical Association*, *86*, 96–107.

Sibley, C. G. (2012). The Mini-IPIP6: Item response analysis of a short measure of the big-six factors of personality in New Zealand. *New Zealand Journal of Psychology*, *41*, 21–31.

Verhelst, N. D., & Glas, C. A. W. (1995). The one parameter logistic model. In G. H. Fischer, & I. W. Molenaar (Eds.), *Rasch Models; Foundations, Recent Developments, and Applications*). New York: Springer Verlag.

von Davier, M., & Carstensen, C. H. (2007). *Multivariate and Mixture Distribution Rasch Models: Extensions and Applications*. New York: Springer Science + Business Media.

World Health Organization (WHO) (2010). *Measuring Health and Disability: Manual for WHO Disability Assessment Schedule*. Geneva: WHO.

How to cite this article: McIntosh J. Evaluating psychological distress data. *Int J Methods Psychiatr Res*. 2017;26:e1551. https://doi.org/10.1002/mpr.1551