

ORIGINAL ARTICLE

# Can Clinicians Use the PHQ-9 to Assess Depression in People with Vision Loss?

Ecosse L. Lamoureux\*, H. Wen Tee<sup>†</sup>, Konrad Pesudovs\*, Julie F. Pallant\*, Jill E. Keeffe\*, and Gwen Rees\*

## ABSTRACT

**Purpose.** To investigate whether the Patient Health Questionnaire-9 (PHQ-9) possesses the essential psychometric characteristics to measure depressive symptoms in people with visual impairment.

**Methods.** The PHQ-9 scale was completed by 103 participants with low vision. These data were then assessed for fit to the Rasch model.

**Results.** The participants' mean  $\pm$  standard deviation (SD) age was  $74.7 \pm 12.2$  years. Almost one half of them ( $n = 46$ ; 44.7%) were considered to have severe vision impairment (presenting visual acuity  $<6/60$  in the better eye). Disordered thresholds were evident initially. Collapsing the two middle categories produced ordered thresholds and fit to the Rasch model ( $\chi^2 = 10.1$ ; degrees of freedom = 9;  $p = 0.34$ ). The mean (SD) items and persons Fit Residual values were  $-0.31$  (1.12) and  $-0.25$  (0.78), respectively, where optimal fit of data to the Rasch model would have a mean = 0 and SD = 1. Unidimensionality was demonstrated confirming the construct validity of the PHQ-9 and there was no evidence of differential item functioning on a number of factors including visual disability. The person separation reliability value was 0.80 indicating that the PHQ-9 has satisfactory precision. There was a degree of mistargeting as expected in this largely non-clinically depressed sample.

**Conclusions.** Our findings demonstrate that the PHQ-9, when scaled with Rasch analysis, forms a linear interval measurement of depressive symptoms suitable for use in a vision impaired population. (Optom Vis Sci 2009;86:139–145)

Key Words: depression, low vision, PHQ-9, Rasch analysis

Visual impairment significantly reduces participation in daily living and increases the need for community and/or family support.<sup>1–3</sup> Individuals with vision impairment are also likely to have poor mental health.<sup>4–6</sup> Indeed, the rates of depression are two to five times greater in older adults with low vision, compared with sighted individuals of similar age.<sup>7</sup> Although a clinical diagnosis of depression can only be made by a qualified health professional, there are several self-report scales available to screen for symptoms of depression such as the Centre for Epidemiological Studies Depression (CES-D),<sup>8</sup> Hospital Anxiety and Depression Scale,<sup>9</sup> and Beck Depression Inventory.<sup>10</sup> With the rela-

tively high rate of depression in the visually impaired population, it has become important to rely on valid screening tools to assess whether visually impaired individuals are suffering from depression.

One screening tool that has been increasingly used is the Patient Health Questionnaire (PHQ).<sup>11</sup> The PHQ was the first self-report questionnaire designed for use in primary care that provides information on specific disorders using the criteria from the Diagnostic and Statistical Manual of Mental Disorders.<sup>12</sup> Its nine-item depression module, the PHQ-9, which is based on the diagnostic criteria of the Diagnostic and Statistical Manual of Mental Disorders for major depression and other depressive disorders, is increasingly being used as a brief diagnostic and severity measure in research and clinical practice.<sup>13</sup> A PHQ-9 score  $\geq 10$  has been shown to have a sensitivity of 88% and a specificity of 88% for major depression.<sup>13</sup> The PHQ-9 offers a number of advantages as it is a relatively short scale to administer (half the length of many other depression measures); has been validated for telephone administration<sup>14</sup>; and scores of 5, 10, 15, and 20 represent valid thresholds demarcating the lower limits of mild, moderate, moderately severe, and severe depression.

Using Rasch analysis, the purpose of this study was to determine whether the PHQ-9 possesses the required psychometric charac-

\*PhD

<sup>†</sup>MBBS

Singapore Eye Research Institute, Singapore National Eye Centre, Singapore (ELL), Centre for Eye Research Australia, University of Melbourne, Melbourne, Victoria, Australia (ELL, WHT, JEK, GR), NH&MRC Centre for Clinical Eye Research, Flinders University, Flinders Medical Centre-Australia, Bedford Park, South Australia, Australia (KP), School of Rural Health, Shepparton, University of Melbourne, Melbourne, Victoria, Australia (JFP), Vision CRC, Sydney, New South Wales, Australia (ELL, JEK), and The Royal Victorian Eye and Ear Hospital, Melbourne, Victoria, Australia (ELL, JEK, GR).

teristics to measure depressive symptoms in people with visual impairment. We specifically assessed the PHQ-9 for unidimensionality, interval scaling, internal consistency reliability, concurrent validity, appropriate targeting of person ability to item difficulty, and the absence of differential item functioning (DIF) for important disease-related factors. Clinically, this information is important as the relatively short PHQ-9 can be used as a time-effective tool to identify patients at risk for depression.

## METHODS

### Participants

Participants were recruited from the outpatient eye clinics at Royal Victorian Eye and Ear Hospital and private eye clinics of Victoria Parade Eye Consultants. The eligibility criteria were presenting visual acuity  $<6/12$  in the better eye with habitual correction; aged 18 years and above; ability to converse in English; adequate hearing (including the use of a hearing aid if necessary); and no cognitive impairment as determined by the 6-Item Cognitive Impairment Test.<sup>15</sup> Eligible individuals who agreed to participate gave their informed consent and were either interviewed in person on the day or at a convenient time. Clinical data were collected and participants provided sociodemographic data, and medical and mental histories. Current participation in low-vision rehabilitation was also recorded. Ethical approval was obtained from the Royal Victorian Eye and Ear Hospital Human Research and Ethics Committee. This research adhered to the tenets of the Declaration of Helsinki.

### The PHQ-9

The nine items of the PHQ-9 scale are shown in Table 1. The wording preceding these items is “Over the last 2 week, how often have you been bothered by any of the following problems?” Responses to the PHQ-9 are rated using a four-category Likert scale: 0 = “Not at all,” 1 = “Several days,” 2 = “More than half the days” (2) and 3 = “Nearly every day.” The total score is calculated by summing each of the PHQ-9 items (range 0 to 27) with higher scores indicating the presence of more symptomatology. The scores of 5, 10, 15, and 20 represent thresholds demarcating the

lower limits of mild, moderate, moderately severe, and severe depression. The criterion, construct, and external validity of the PHQ-9 have been well established using large samples from a range of patient populations.<sup>16,17</sup>

### Rasch Analysis

In psychometric terms, the Rasch measurement model, also known as Rasch analysis, is a mathematical description of the way in which respondents must interact with test items to produce linear measurement. Rasch analysis states that the probability of an individual’s choosing a response on a particular item depends on both the person ability and item difficulty. Thus, Rasch analysis is taken as a criterion for the structure of the responses which they should be satisfied rather than a simple statistical description of the responses.<sup>18–21</sup>

Rasch analysis was undertaken to determine specific aspects of validity, reliability, and measurement of the PHQ-9. We used Rasch Unidimensional Measurement Models (RUMM) 2020 (RUMM 2020. Perth: RUMM Laboratory, 2003) and the Andrich single rating scale to determine whether the data fitted the Rasch model.<sup>22</sup> Content validity was evaluated using person and item Fit Residual statistics, where it is expected that the mean and SD (standard deviation) values approximate 0 and 1, respectively. An overall item-trait interaction score ( $\chi^2$ ) with a statistically non-significant probability value ( $p > 0.05$ ) indicates fit to the model and that hierarchical ordering of the items (i.e., from difficult to easy) is consistent across all levels of depressive symptoms. An estimate of the internal consistency reliability of the scale based on the person separation reliability (PSR) was also reported.<sup>23</sup> PSR values (similar to Cronbach alpha) of  $\geq 0.7$  are suitable for group use.<sup>24</sup> This is closely linked to the targeting of the scale as it differentiates the number of statistically distinct groups of respondents that can be identified on the trait.<sup>25</sup>

We also assessed disordered thresholds which occur when participants have difficulty discriminating between the response options. This means literally that a category expected to be “harder” than an adjacent category was actually “easier,” but often represents interchangeability of categories. Disordered thresholds indicate a lack of the invariance of items, which is the ratio of difficulties between any pair of items, remains constant across the ability levels of respondents. Category collapsing is often the solution to disordered thresholds.

DIF testing was undertaken to determine whether different groups within the sample despite equal levels of functioning, respond differently to individual items. In RUMM 2020, Uniform DIF can occur when one group shows a consistent systematic difference in their responses to an item, across the whole range of the attribute being measured. When there is non-uniformity in the differences between the groups (e.g., it varies across levels of the attribute) then this is referred to as Non-Uniform DIF. In RUMM, the presence of DIF can be detected both statistically and graphically. Analysis of variance is conducted for each item comparing scores across each level of the “person factor” and across different levels of trait (referred to as class intervals). Uniform DIF is indicated by a significant main effect for the person factor, whereas the presence of non-Uniform DIF is indicated by a significant interaction effect (person factor  $\times$  class interval). The

**TABLE 1.**

The nine items of the PHQ-9 depression screening scale

1. Little interest or pleasure in doing things
2. Feeling down, depressed, or hopeless
3. Trouble falling or staying asleep or sleeping too much
4. Feeling tired or having little energy
5. Poor appetite or overeating
6. Feeling bad about yourself—or that you are a failure or have let yourself or your family down
7. Trouble concentrating on things, such as reading the newspaper or watching television
8. Moving or speaking so slowly that other people could have noticed. Or the opposite—being so fidgety or restless that you have been moving around a lot more than usual
9. Thoughts that you would be better off dead or of hurting yourself in some way

Bonferroni adjusted *p* value used for this set of analyses was 0.05/9 = 0.006. In this article, Uniform and Non-Uniform DIFs were assessed for age (<60, 60 to 69, 70 to 79, ≥80); gender (male, female); life events (yes, no); degree of visual impairment (<6/12 to 6/18, <6/18 to 6/60, <6/60 to 3/60, <3/60); ocular conditions (age-related macular degeneration, diabetic retinopathy, glaucoma, other); non-ocular comorbidity (yes, no); and levels of visual disability assessed by the Impact of Vision Impairment (IVI) scale<sup>26,27</sup> (low, middle, and high tertiles).

Finally, the unidimensionality of the scale is assessed using a principal components analysis (PCA) of the residuals. This allows for a test of the local independence of the items. This test implies that once the Rasch factor has been taken into account, there should be no further associations between the items other than random associations. In RUMM 2020, the unidimensionality of the PHQ-9 was assessed using PCA of the residuals. Unidimensionality took the form of an independent *t*-test technique looking for deviations between two sets of person locations, on a person by person basis.<sup>28,29</sup> The two location sets were derived from two subsets of items (“positive” and “negative” loadings subsets) from the final scale and identified by the loadings of item residuals in a PCA. The overall proportion of *t* values falling outside a ±1.96 range should be ≤5% to confirm unidimensionality.

Unidimensionality, in addition with the other adequate fit statistics of the data to the Rasch model, would support the specific validity, reliability, and measurement characteristics of the PHQ-9.<sup>30</sup> To maximize the retention of the initial character of the PHQ-9, a minimalist approach was taken to item and scale changes such that only those changes essential for satisfactory scale functioning were made.

## Centre for Epidemiological Studies Depression

We used the CES-D scale as a measure of concurrent validity as it has been used previously to assess depressive symptoms among the vision impaired.<sup>8,31–34</sup> The CES-D is a 20-item instrument that measures current levels of depressive symptoms. The CES-D has been found to be a highly reliable measure and its validity as a measure of depression has been supported in community studies of adults.<sup>8,35</sup>

## The Impact of Vision Impairment Questionnaire

The IVI, a 28-item vision-specific quality of life questionnaire,<sup>26,27</sup> was included in the study to determine whether the PHQ-9 can effectively measure depressive symptomatology independent of visual disability.

## Statistical Analyses

Descriptive statistical analyses were performed to characterize the sociodemographic and clinical data of the participants using the SPSS statistical software (Version 15.0. SPSS Science, Chicago, IL).

## RESULTS

### Participants' Characteristics

The mean age of the 103 participants was 74.7 years (Table 2). The majority were female (*n* = 62; 61.2%) and had age-related

**TABLE 2.**

The participants' characteristics (N = 103)

	Mean ± SD (range)	
Age (yr)	74.7 ± 12.2 (23–95)	
Gender	Men	41 (39.8%)
Life events (last 12 months)	Negative	15 (14.6%)
Presenting visual acuity (better eye)	<6/12 to 6/18	13 (12.6%)
	<6/18 to 6/60	44 (42.7%)
	<6/60–3/60	11 (10.7%)
	<3/60	35 (34.0%)
Main cause of vision loss	AMD	56 (54.4%)
	Diabetic retinopathy	29 (28.3%)
	Glaucoma	9 (8.7%)
	Other	9 (8.8%)
Duration of vision impairment (yr)	Median (min, max)	4 (0.06, 70)
Comorbidity	Yes	80 (77.7%)

AMD, age-related macular degeneration.

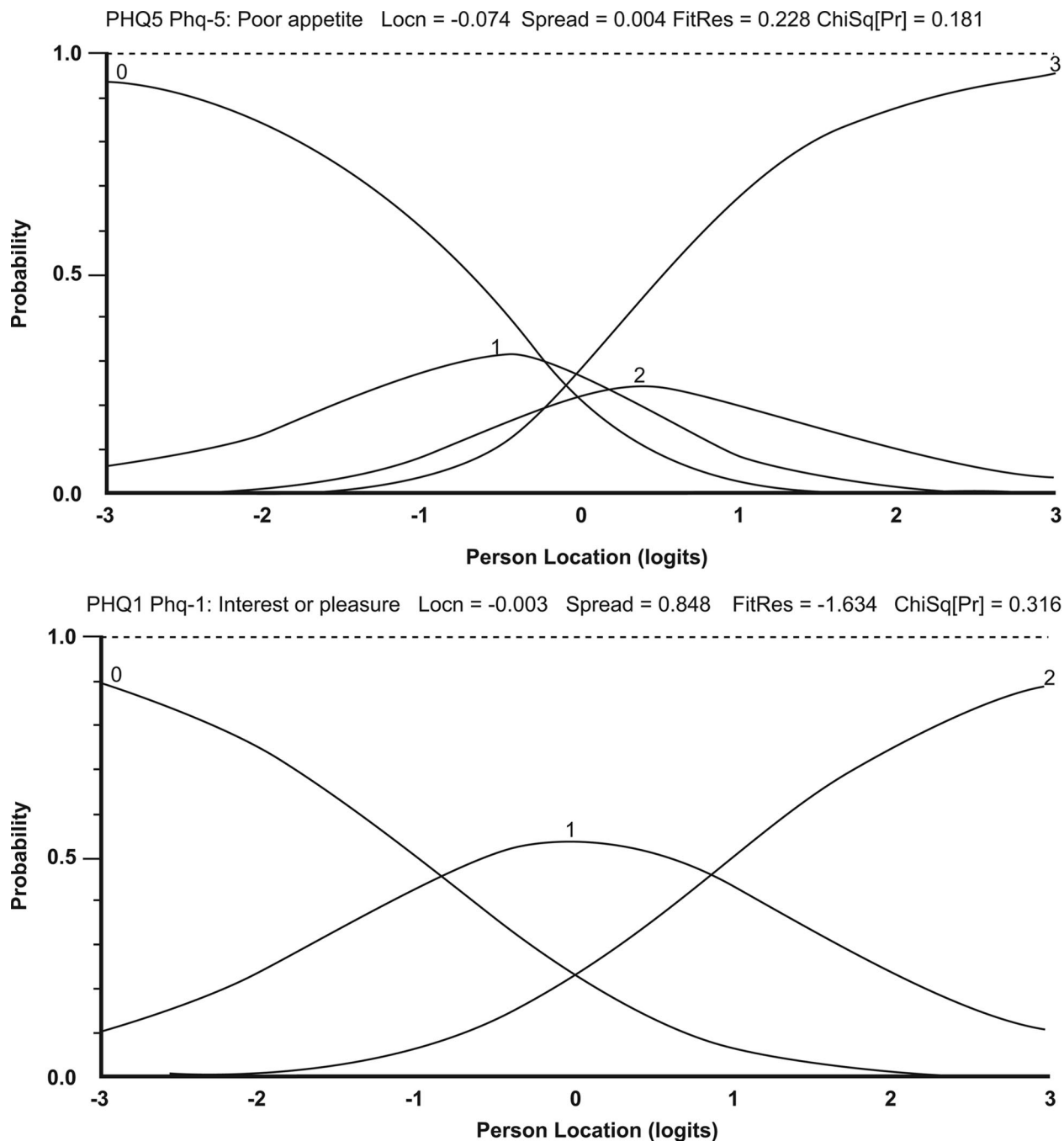
macular degeneration (*n* = 56; 54.4%). Almost one half of them (*n* = 46; 44.7%) were considered to have severe vision impairment (presenting visual acuity <6/60 in the better eye). Fifteen of the participants (14.6%) recorded PHQ-9 raw scores ≥10 suggesting risks of moderate to severe depression.

## Rasch Analysis of the PHQ-9

The data of the PHQ-9 were initially fitted to the Rasch model. There was evidence of disordered thresholds for all items. Disordered thresholds is a sign that the categories are not working as intended and can occur when there are too many response options, or when the labeling of options is similar to each other, potentially confusing or open to misinterpretation. For example, as can be seen from Fig. 1 (left panel), the response category 2 for one of the nine items “Poor appetite or overeating” was not consistently chosen by the participants (i.e., that the difficulty of a higher threshold was lower than that of its adjacent lower threshold), indicating that category 2 was never the most likely response of any participant and was thus redundant. This necessitated category 2 to be collapsed with either category 1 or 3. Both proved equally effective in producing ordered thresholds for all items. We arbitrarily chose collapsing categories 1 and 2 (Fig. 1, right panel).

After category ordering, the PHQ-9 showed fit to the Rasch model with a non-significant Item-Trait Interaction total  $\chi^2$  probability value ( $\chi^2 = 10.1$ , degrees of freedom = 9, *p* = 0.34). The mean (SD) items and persons Fit Residual values were −0.31 (1.12) and −0.25 (0.78), respectively, where optimal fit of data to the model would have a mean = 0 and SD = 1. All items showed Fit Residuals values <2.5 with probability scores >0.05 indicating no significant deviation from the model (Table 3). The PSR value was 0.80 and demonstrates substantial internal consistency reliability.

The person-item location map (Fig. 2) displays the participants' scores on the Rasch calibrated scale (on the left hand side) and levels of symptomatology of each PHQ-9 item on the right hand side. Participants with high levels of depressive symptoms and items with high levels of symptomatology are at the top of the diagram. As anticipated, there was an uneven spread of items across



**FIGURE 1.**

Upper panel: category probability curves showing “disordered thresholds” for the item “Poor appetite or overeating.” The response category 2 does not have a range along the ability scale where it is most likely to be chosen. Lower panel: category probability curves showing ordered thresholds after having collapsed categories 1 and 2.

the full range of the participant’s scores, which indicates that most of the participants had low levels of depressive symptoms.

All PHQ-9 items were found to be free from DIF for age, gender, the duration of visual impairment, type and degree of visual impairment and comorbidity, and visual disability as measured by IVI with probability values exceeding the alpha value for each of the person factors assessed ( $p > 0.05$ ). Table 4 shows the results for the PHQ-9 items for Uniform and Non-Uniform DIFs for one person factor, i.e., level of visual impairment. This finding indicates that the PHQ-9 is a valid scale to assess depressive symptoms across the various categories of each of these variables.

PCA of the residuals identified two subsets of items consisting of the highest positive (PHQ-4 and PHQ-5) and negative loading items (PHQ-1, 2, and 8). Only 2.5% of estimates were found to be significantly different for these participants. These values are less than the recommended cut point of 5% and therefore no evidence of multidimensionality was detected. This finding confirms the internal construct validity of the PHQ-9 and that the instrument is measuring one underlying trait (level of depressive symptoms) that it purports to measure.

These results collectively show that the PHQ-9 is a unidimensional scale and the overall score has estimates of an interval scaling.



**TABLE 3.**

Category frequencies and Fit indices (location, standard error, fit residuals,  $\chi^2$ , and probability values) of the PHQ-9 items to the Rasch model before and after rescoring

Before rescoring	Category response frequencies				Location	SE	FitResid	$\chi^2$	Prob
	1	2	3	4					
PHQ-1	51	19	7	3	-0.04	0.15	-1.11	0.62	0.43
PHQ-2	30	35	8	7	-0.64	0.13	-1.66	1.27	0.26
PHQ-3	31	18	17	14	-0.89	0.12	1.92	0.07	0.79
PHQ-4	22	22	10	26	-1.29	0.12	-0.35	0.31	0.58
PHQ-5	56	9	8	7	-0.07	0.15	0.23	1.79	0.18
PHQ-6	58	13	6	3	0.25	0.17	-0.93	3.84	0.05
PHQ-7	57	14	4	5	0.12	0.16	0.20	0.39	0.53
PHQ-8	71	6	2	1	1.13	0.26	-0.30	2.34	0.13
PHQ-9	72	6	2	0	1.45	0.31	-0.86	2.00	0.16

After rescoring	Category response frequencies				Location	SE	FitResid	$\chi^2$	Prob
	1	2	3	4					
PHQ-1	51	26	3	0.00	0.22	-1.63	1.01	0.32	
PHQ-2	30	43	7	-0.98	0.19	-1.25	0.13	0.72	
PHQ-3	31	35	14	-1.17	0.19	2.03	1.36	0.24	
PHQ-4	22	32	26	-1.92	0.18	0.13	0.53	0.47	
PHQ-5	56	17	7	0.05	0.22	0.24	0.82	0.37	
PHQ-6	58	19	3	0.41	0.24	-1.14	3.21	0.07	
PHQ-7	57	18	5	0.18	0.23	0.60	0.34	0.56	
PHQ-8	71	8	1	1.60	0.34	-0.80	1.56	0.21	
PHQ-9	72	8	0	1.83	0.37	-1.00	1.19	0.28	

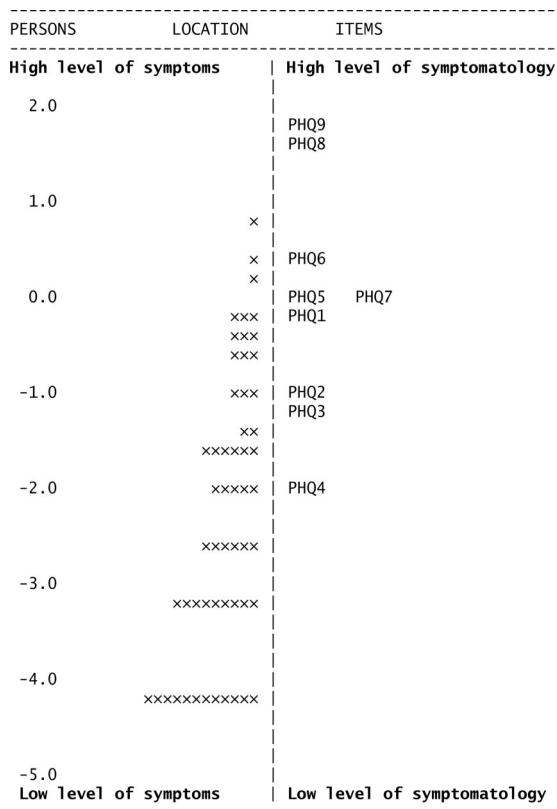
All items showed Fit Residuals values <2.5 and probability scores >0.05.

SE, standard error; FitResid, Fit Residuals;  $\chi^2$ , chi square; Prob, probability values.

The negative mean person location logit value (-2.21) indicates that the participants on average have a substantially lower levels of depressive symptoms than that tapped by the PHQ-9 items (0 logits). The two items with the highest mean levels of symptomatology were PHQ-9 “Thoughts that you would be better off dead or of hurting yourself in some way” and PHQ-8 “Moving or speaking so slowly that other people could have noticed” (1.83 and 1.60 logits, respectively). Conversely, the two items with the lowest levels of depressive symptomatology were PHQ-4 “Feeling tired or little energy” and PHQ-3 “Trouble falling or staying asleep or sleeping too much” with logit scores of -1.92 and -1.17, respectively.

**Concurrent Validity and Proposed Cutoff Score on the Rasch Calibrated Scale**

Using the raw scores and original categorization, the commonly used cutoff on the CES-D ( $\geq 16$ ) identified 19.4% (n = 20) of the respondents as having risk for clinical depression. Similarly, using the PHQ-9 raw summary scores, 14.6% (n = 15) of the participants recorded scores  $\geq 10$  suggesting risks of moderate to severe depression. All of the 15 cases of moderate to severe depression



**FIGURE 2.** Person-Item location map of the Rasch-scaled PHQ-9 (showing the distribution of calibrated participants’ scores (left hand side) and item locations (right-hand side)).

**TABLE 4.**

The results for the PHQ-9 items for Uniform DIF (middle section headed “Level of visual impairment”) and Non-Uniform DIF (headed Class interval by “Level of visual impairment”)

Item	Class interval			Level of visual impairment			Class interval × level of visual impairment		
	MS	F	Prob	MS	F	Prob	MS	F	Prob
PHQ1	1.61	1.58	0.20	0.11	0.11	0.74	0.65	0.64	0.59
PHQ2	4.54	3.48	0.02	0.39	0.30	0.59	4.96	3.79	0.01
PHQ3	1.12	1.26	0.29	0.02	0.03	0.87	0.46	0.52	0.67
PHQ4	1.42	1.63	0.19	0.04	0.04	0.83	0.06	0.07	0.98
PHQ5	3.65	3.37	0.01	0.12	0.18	0.67	0.20	0.29	0.83
PHQ6	0.29	0.30	0.83	0.44	0.46	0.50	1.91	1.99	0.12
PHQ7	0.84	0.90	0.45	0.06	0.07	0.80	0.87	0.93	0.43
PHQ8	0.65	0.51	0.68	0.68	0.54	0.47	1.05	0.83	0.48
PHQ9	0.41	0.34	0.80	0.28	0.23	0.63	1.03	0.85	0.47

The Bonferroni adjusted p value was 0.006. MS, mean-square.

were identified by both depression scales. A Rasch-scaled PHQ-9 cutoff score of  $> -0.6$  logits was found to effectively identify the same 15 cases of moderate to severe depression identified using the traditional cutoff scores  $\geq 10$ . Similarly, our data indicate that scores of -1.23, 0.29, 1.99, and  $>4.1$  logits represent thresholds

demarcating the lower limits of mild, moderate, moderately severe, and severe depression, respectively.

## DISCUSSION

We used Rasch analysis to assess whether the PHQ-9 possesses the essential psychometric characteristics to effectively assess depressive symptoms in people with visual impairment. We found that the PHQ-9 meet the formal requirements of measurement as defined by the Rasch model following response category collapsing. The scale possesses demonstrated validity (construct and concurrent), reliability (internal consistency), and unidimensionality. The PHQ-9 also showed no evidence of DIF or misfitting items to the overall latent trait. These findings suggest that with the overall transformation to interval scaling, the PHQ-9 has the measurement and psychometric properties to be an effective screening tool for assessing depressive symptoms in visually impaired individuals.

To our knowledge, this is the first time that the PHQ-9 has undergone Rasch analysis in people with visual impairment making it difficult to compare our findings. In cancer patients, the PHQ-9 and seven other scales were used to develop an item bank for assessing and screening for psychological distress in cancer patients.<sup>36</sup> Only one item of the PHQ-9 was found to misfit (poor appetite or overeating) probably because of the fact that it was not specific to psychological distress in cancer patients. This item did not misfit in our analysis. The only revision necessary was collapsing the two middle response categories. This is because the categories “Several days” and “More than half the days” are used interchangeably. It is very likely that the labeling of options being similar to one another may have potentially led to confusion or misinterpretation by our respondents. It is equally likely that three levels may adequately classify people’s responses to each item. Further validation studies are, however, needed to confirm the three-response scale with a suitably named middle category.

Although collapsing response categories in this way represents an important improvement in scale functioning, it has serious implications for use of the PHQ-9 as screening instruments. When applied for screening, the diagnoses are made by the number of symptoms people rate as experiencing. So collapsing categories changes the original screening cutoff points. Our results cannot determine the impact of this change or the validity of a new interpretation although it is likely that the sensitivity/specificity will improve as category collapsing will have reduced noise in the measurement, which typically improves precision in instrument application.<sup>18,19</sup> We have identified the new cutoff point on a Rasch-calibrated overall score to be greater than  $-0.6$  logits to effectively identify people with moderate to severe depression symptoms. We have also provided the Rasch scores for the limits of mild, moderate, moderately severe, and severe depression. Further work is, however, needed with the modified response scale and new cut points to identify appropriate levels of symptoms for depression need to be established against other validated measurements.

In addition to using a modern, sophisticated statistical approach such as Rasch analysis to test the psychometric properties of the PHQ-9, we also assessed whether the scale was effectively measuring depressive symptomatology independent of visual disability. We wanted to assess whether bias exists for an item among subgroups in the sample particularly those with

differential levels of vision-specific quality of life. We showed the PHQ-9 items were free from DIF for the restriction of participation (mild, moderate, or severe) as assessed by the IVI. This finding indicates that the PHQ-9 can assess depressive symptoms in visually impaired people independent of their visual disability and demonstrate its appropriateness to be used in this population.

One of the potential limitations of the current study is the relatively small sample as this may have affected the assessment for DIF undertaken in this study. To account for this limitation, we did not apply a Bonferroni adjusted p-value for our DIF tests (0.005). Rather, we looked for substantial trends indicative of differential with a p-value set at 0.05. There was still no evidence of DIF on all our person factors although our findings should be further tested in a larger visually impaired sample.

In conclusion, reengineering scales using Rasch analysis is consistent with the Task Force for Statistical Inference in psychology journals recommendation that the use of Item-Response Test theory for assessment of the psychometric properties of scales is needed.<sup>37</sup> We adhered to that suggestion for the PHQ-9 and it was found to possess internal construct validity through unidimensionality, invariance of items through interval-level scaling; appropriate category ordering, and the absence of bias among subgroups in the sample. Collectively, these findings suggest that the PHQ-9 has the potential to be used as a tool to measure depressive symptoms in people with visual impairment. However, the authors would caution that their findings are based on a relatively small sample and need to be replicated using a larger sample.

## ACKNOWLEDGMENTS

*Received May 29, 2008; accepted July 29, 2008.*

## REFERENCES

1. Wang JJ, Mitchell P, Smith W, Leeder SR. Factors associated with use of community support services in an older Australian population. *Aust N Z J Public Health* 1999;23:147–53.
2. Lamoureux EL, Hassell JB, Keeffe JE. The determinants of participation in activities of daily living in people with impaired vision. *Am J Ophthalmol* 2004;137:265–70.
3. Wang JJ, Mitchell P, Smith W, Cumming RG, Leeder SR. Incidence of nursing home placement in a defined community. *Med J Aust* 2001;174:271–5.
4. Rovner BW, Zisselman PM, Shmueli-Dulitzki Y. Depression and disability in older people with impaired vision: a follow-up study. *J Am Geriatr Soc* 1996;44:181–4.
5. Rovner BW, Casten RJ, Hegel MT, Tasman WS. Minimal depression and vision function in age-related macular degeneration. *Ophthalmology* 2006;113:1743–7.
6. Black SA. Increased health burden associated with comorbid depression in older diabetic Mexican Americans. Results from the Hispanic Established Population for the Epidemiologic Study of the Elderly survey. *Diabetes Care* 1999;22:56–64.
7. Evans JR, Fletcher AE, Wormald RP. Depression and anxiety in visually impaired older people. *Ophthalmology* 2007;114:283–8.
8. Radloff LS. The CES-D Scale: a self-report depression scale for research in the general population. *Appl Psychol Meas* 1977;1:385–401.

9. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983;67:361–70.
10. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry* 1961;4:561–71.
11. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary care evaluation of mental disorders. Patient Health Questionnaire. *JAMA* 1999;282:1737–44.
12. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR, 4th ed. Text revision. Washington, DC: American Psychiatric Association; 2000.
13. Koneke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001;16:606–13.
14. Pinto-Meza A, Serrano-Blanco A, Penarrubia MT, Blanco E, Haro JM. Assessing depression in primary care with the PHQ-9: can it be carried out over the telephone? *J Gen Intern Med* 2005;20:738–42.
15. Katzman R, Brown T, Fuld P, Peck A, Schechter R, Schimmel H. Validation of a short orientation-memory-concentration test of cognitive impairment. *Am J Psychiatry* 1983;140:734–9.
16. Lowe B, Spitzer RL, Grafe K, Koneke K, Quenter A, Zipfel S, Buchholz C, Witte S, Herzog W. Comparative validity of three screening questionnaires for DSM-IV depressive disorders and physicians' diagnoses. *J Affect Disord* 2004;78:131–40.
17. Martin A, Rief W, Klaiberg A, Braehler E. Validity of the Brief Patient Health Questionnaire Mood Scale (PHQ-9) in the general population. *Gen Hosp Psychiatry* 2006;28:71–7.
18. Garamendi E, Pesudovs K, Stevens MJ, Elliott DB. The refractive status and vision profile: evaluation of psychometric properties and comparison of Rasch and summated Likert-scaling. *Vision Res* 2006;46:1375–83.
19. Norquist JM, Fitzpatrick R, Dawson J, Jenkinson C. Comparing alternative Rasch-based methods vs raw scores in measuring change in health. *Med Care* 2004;42:125–36.
20. Pesudovs K. Patient-centred measurement in ophthalmology—a paradigm shift. *BMC Ophthalmol* 2006;6:25.
21. Pesudovs K, Garamendi E, Elliott DB. The Quality of Life Impact of Refractive Correction (QIRC) Questionnaire: development and validation. *Optom Vis Sci* 2004;81:769–77.
22. Andrich D. Rating formulation for ordered response categories. *Psychometrika* 1978;43:561–73.
23. Mallinson T, Stelmack J, Velozo C. A comparison of the separation ratio and coefficient alpha in the creation of minimum item sets. *Med Care* 2004;42:117–24.
24. Nunnally CJ. *Psychometric Theory*. New York: McGraw-Hill; 1978.
25. Fisher WP Jr. Reliability statistics. Rasch Measurement: Rasch Meas Trans 1992;6:238. Available at: <http://www.rasch.org/rmt/rmt63i.htm>. Accessed November 5, 2008.
26. Lamoureux EL, Pallant JF, Pesudovs K, Hassell JB, Keeffe JE. The impact of Vision Impairment Questionnaire: an evaluation of its measurement properties using Rasch analysis. *Invest Ophthalmol Vis Sci* 2006;47:4732–41.
27. Lamoureux EL, Pallant JF, Pesudovs K, Rees G, Hassell JB, Keeffe JE. The impact of vision impairment questionnaire: an assessment of its domain structure using confirmatory factor analysis and rasch analysis. *Invest Ophthalmol Vis Sci* 2007;48:1001–6.
28. Smith EV Jr. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J Appl Meas* 2002;3:205–31.
29. Tennant A, Pallant JF. Unidimensionality matters (A tale of two Smiths?). *Rasch Meas Trans* 2006;20: 1048–51. Available at: <http://www.rasch.org/rmt/rmt201c.htm>. Accessed November 5, 2008.
30. Lundgren-Nilsson A, Grimby G, Ring H, Tesio L, Lawton G, Slade A, Penta M, Tripolski M, Biering-Sorensen F, Carter J, Marincek C, Phillips S, Simone A, Tennant A. Cross-cultural validity of functional independence measure items in stroke: a study using Rasch analysis. *J Rehabil Med* 2005;37:23–31.
31. Horowitz A, Reinhardt JP, Boerner K. The effect of rehabilitation on depression among visually disabled older adults. *Aging Ment Health* 2005;9:563–70.
32. Rovner BW, Casten RJ, Tasman WS. Effect of depression on vision function in age-related macular degeneration. *Arch Ophthalmol* 2002;120:1041–4.
33. Horowitz A, Reinhardt JP, Kennedy GJ. Major and subthreshold depression among older adults seeking vision rehabilitation services. *Am J Geriatr Psychiatry* 2005;13:180–7.
34. Rovner BW, Casten RJ. Activity loss and depression in age-related macular degeneration. *Am J Geriatr Psychiatry* 2002;10:305–10.
35. Aneshensel CS, Clark VA, Frerichs RR. Race, ethnicity, and depression: a confirmatory analysis. *J Pers Soc Psychol* 1983;44: 385–98.
36. Smith AB, Rush R, Velikova G, Wall L, Wright EP, Stark D, Selby P, Sharpe M. The initial development of an item bank to assess and screen for psychological distress in cancer patients. *Psychooncology* 2007;16:724–32.
37. Wilkinson L. Task force on Statistical Inference. Statistical methods in psychology journals: guidelines and explanations. *Am Psychol* 1999;54:594–604.

**Ecosse Lamoureux**

*Department of Ophthalmology  
Centre for Eye Research Australia  
University of Melbourne  
32 Gisborne St., East Melbourne  
Victoria 3002, Australia  
e-mail: ecosse@unimelb.edu.au*