PEER REVIEW HISTORY

This paper was submitted to a another journal from BMJ but declined for publication following peer review. The authors addressed the reviewers' comments and submitted the revised paper to BMJ Open. The paper was subsequently accepted for publication at BMJ Open.

(This paper received three reviews from its previous journal but only two reviewers agreed to published their review.)

ARTICLE DETAILS

TITLE (PROVISIONAL)	The psychometric properties of the Norwegian version of the short form of The Problem Areas in Diabetes scale (PAID-5) – a validation study
AUTHORS	Vislapuu, Maarja; Broström, Anders; Igland, Jannicke; Vorderstrasse, A; Iversen, Marjolein M

VERSION 1 – REVIEW

REVIEWER	Bik-Wai Bilvick Tai Caritas Institute of Higher Education Hong Kong, China
REVIEW RETURNED	23-Apr-2018

GENERAL COMMENTS	Thank you for the opportunity to review this manuscript. This manuscript describes a test-retest validation study that examined the psychometric properties of PAID-5 conducted in Norway. In general, this manuscript is well-structured with specific subheadings. The transition of thoughts runs smoothly that is easy to read and review. The background section provides a good overview of what is known in the literature, and the rationale to conduct the study that corresponds to the existing knowledge gap. However, modifications and clarifications in different sections are opportunities for this manuscript before it can be considered for publication. Each section and subheading in the manuscript has been reviewed with my comments and suggestions listed below: General: 1. Much of the content throughout the manuscript can be written in a more precise and concise approach. Introduction/background: 1. Too much content in the introduction and background section. Suggest to combine the two sections into one section, and to only keep the most important content relevant to this validation study. E.g. take away "Diabetes-specific distress should not be confused with general emotional distress and it is also conceptually distinct from major depressive disorder [11]." 2. More information on type 1 and 2 diabetes (e.g. prevalence) in Norway shall be provided in the Introduction section. 3. P.5 L25-26: Please state in which country the original PAID-20 was developed.

	Methods: 1. P.6 L45-46: Please explain why only DM patients diagnosed more than 1 year ago were included in this study. 2. P.6 L54-56: Please clarify how many surveys and/or survey questions were completed by the participants in this study, as the authors mentioned: “Patients meeting the inclusion criteria received a paper questionnaire by mail…”. Did the “paper questionnaire” contain questions from more than one questionnaire (e.g. PAID-5, PAID-20). 3. Was WHO-5 validated in Norway previously to be used in this study? 4. P.8 L19-21: Please remove the sub-section on “Patient and Public Involvement”. 5. P.8 L42-43: Please mention how the strength of Pearson’s correlations was categorized. 6. P.9 L5-6: PAID-20 was already validated in Norway in a previous study (Graue M et al. International Journal of Nursing Studies 49 (2012) 174–182). Please justify why the internal consistency of PAID-20 was measured again in the current study. Results: 1. The PAID-5 score in this current study should first be reported in the results but not the discussion section. 2. P.10 L40-44: “There were no significant differences between the mean scores of PAID-5 for the three different treatment groups: insulin (n=87), oral medication (n=28) or both (n=27) (p=0.90).” Please discuss this result in the discussion section. Discussion: 1. P.12 L26-27: Please standardize the chemistry unit of HbA1c in the text and Table 2. 2. P.13 L7-23: I do not see this paragraph fits well into the discussion section of a validation study article. Instead, the authors may like to mention the strengths of the current study in the manuscript 3. P.13 L28-34: Please check the grammar and writing style of these sentences to ensure the content is appropriately delivered. References: 1. Please format Reference 27 as some characters are not displayed properly. 2. Unbold “179” for Reference 35. Table 2: 1. Please check again n (%) and mean ± SD are used properly for each parameter.

REVIEWER	AHP Iran
REVIEW RETURNED	11-May-2018

GENERAL COMMENTS	I reviewed the paper with interested. However, I think that a simple validation study can not add anything to the current literature. Several versions of the PAID have been published before. Therefore, the paper should provide a strong justification for the novelty. Here are my comments: Abstract: please specify time interval for test-retest reliability. I would see factorial invariance across gender or presence

	complication and Rasch model for this scale. It would be great if the authors calculate average variance extracted and composite reliability. I am not really sure how the authors want to generalize their consecutive sample size into the national. I would omit or revise the last sentence of conclusion Introduction: I don't agree to separate introduction with background. It would be great to add some sentences on why a measure should adopt from a culture into another culture. Method: the study is methodological not really test retest! Sample size justification is not really true from the Wolf's study. They said that a range of sample size requirements (i.e., from 30 to 460 cases), meaningful patterns of association between parameters and sample size. Typically 5 to 20 respondents/parameter estimated per each parameters (Bentler, 1995). I cannot understand what is mean for Patients and or public! Method of estimation for the CFA should be added. How the authors dealt with missing data in the CFA? Results: please provide all factor loadings for the CFA model. Table 1 should be omitted. Table 4 could be reported in the text. The authrs should discuss why they added an covariance error terms between items 3 and 16. Brown (2015) highligts that you need a good justification to add correlated errors between some indicators of your construct and these correlation, that you should not only model them to reach the common cut-offs for good model fit and to be consistent with the rule you apply. He presents several potential reasons why such correlated errors can occur, like shared method variance due to different wording compared to other indicators, or specific item content (even you might not had this assumption a prior, I think you can defend them, if such reason is plausible). I can really recommend his chapter: Brown, T. A. (2015). Confirmatory factor analysis for applied research. Guilford Publications. Discussion : it is good!

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

General: Much of the content throughout the manuscript can be written in a more precise and concise approach.

We have edited the manuscript and tried to make it more precise and concise.

Introduction/background:

1. Too much content in the introduction and background section. Suggest to combine the two sections into one section, and to only keep the most important content relevant to this validation study. E.g. take away "Diabetes-specific distress should not be confused with general emotional distress and it is also conceptually distinct from major depressive disorder [11]."

As suggested by the reviewers we have combined the introduction and background section and only kept the most relevant content for this validation study. The total number of words in the new "introduction" is reduced with 201 words as requested.

2. More information on type 1 and 2 diabetes (e.g. prevalence) in Norway shall be provided in the Introduction section.

We have added detailed information on prevalence and incidence of type 1 and type 2 diabetes in Norway in the introduction section (lines 3-7/ page 4).

3. P.5 L25-26: Please state in which country the original PAID-20 was developed.

We have included information as to which country the original PAID-20 was developed in (line 2/para4/ page 4).

Methods:

1. P.6 L45-46: Please explain why only DM patients diagnosed more than 1 year ago were included in this study.

As suggested we have added information and explained why we only included DM patients diagnosed for more than 1 year ago in the study (lines 14-16/ para 1/ page 6).

2. P.6 L54-56: Please clarify how many surveys and/or survey questions were completed by the participants in this study, as the authors mentioned: "Patients meeting the inclusion criteria received a paper questionnaire by mail…". Did the "paper questionnaire" contain questions from more than one questionnaire (e.g. PAID-5, PAID-20).

We agree that this was unclear and have clarified the paragraph (lines 1-4/ para 2/ page 6). The questionnaire included background questions, one overall perceived health question, the WHO-5, HADS, HFSII and PAID questionnaire. In total 68 questions as described in the measurement section (lines 158-194/ page6-7).

3. Was WHO-5 validated in Norway previously to be used in this study?

We did not find a relevant validation study in Norway. Therefore, we added a systematic review of the literature on the WHO-5. The study shows that the scale has adequate validity both as a screening tool for depression and as an outcome measure in clinical trials and has been applied successfully across a wide range of study fields (lines 6/ para 4/ page 7).

4. P.8 L19-21: Please remove the sub-section on "Patient and Public Involvement".

We have removed the section as suggested.

5. P.8 L42-43: Please mention how the strength of Pearson's correlations was categorized.

When we use the terms moderate and weak in relation to correlation coefficients on page 9 in the results section we have applied the cutoffs as stated in research references, among others, Polit, D. F., & Beck, C. T. (2012): 0-0.19 is regarded as very weak, 0.2-0.39 as weak, 0.40-0.59 as moderate, 0.6-0.79 as strong and 0.8-1 as very strong correlation. We have now added this information in the methods section (lines 5-6/ para 2/ page 8).

6. P.9 L5-6: PAID-20 was already validated in Norway in a previous study (Graue M et al. International Journal of Nursing Studies 49 (2012) 174–182). Please justify why the internal consistency of PAID-20 was measured again in the current study.

We thank the reviewer for this point and we have omitted the text in the methods section.

Results:

1. The PAID-5 score in this current study should first be reported in the results but not the discussion section.

We apologize and have now added the score in the results section (lines 8-9/ para 2/ page 9).

2. P.10 L40-44: "There were no significant differences between the mean scores of PAID-5 for the three different treatment groups: insulin (n=87), oral medication (n=28) or both (n=27) (p=0.90)." Please discuss this result in the discussion section.

The results regarding "no significant differences between the mean scores of the PAID-5 for the three treatment groups" are now addressed in the discussion section (lines 1-8/ para1/ page 12).

Discussion:

1. P.12 L26-27: Please standardize the chemistry unit of HbA1c in the text and Table 2.

We have standardized the chemistry unit of HbA1c in the text and Table 2 accordingly (line 3-4/ para 2/ page 9) (line 3/ para2/ page 12).

2. P.13 L7-23: I do not see this paragraph fits well into the discussion section of a validation study article. Instead, the authors may like to mention the strengths of the current study in the manuscript

We have adjusted the text and have added strengths of the study in the manuscript and adjusted the heading (line 1-3 / para 2/page 13).

3. P.13 L28-34: Please check the grammar and writing style of these sentences to ensure the content is appropriately delivered.

We have clarified the sentences P.13 L28-34 accordingly (lines 5-6/ para 2/ page 13).

References:

1. Please format Reference 27 as some characters are not displayed properly.

2. Unbold "179" for Reference 35.

Thank you. We have made the format changes as suggested.

Table 2:

1. Please check again n (%) and mean ± SD are used properly for each parameter.

We have checked and corrected Table 2**.**

Reviewer: 2

I reviewed the paper with interested. However, I think that a simple validation study can not add anything to the current literature. Several versions of the PAID have been published before. Therefore, the paper should provide a strong justification for the novelty.

The PAID-5 has recently become more widely used in research, because of its brevity. However, there is still limited knowledge of the psychometric properties of the PAID-5 scale in Europe, in particular the factor structure of the PAID-5. As there is a need for a short diabetes distress questionnaire in the Norwegian Diabetes Registry for Adults as well as in Norwegian population based cohort studies this needs to be explored. We have expanded information regarding justification for the study in the introduction section (lines 10-19/ para 1/ page 5).

Abstract: please specify time interval for test-retest reliability.

We have added time interval for test-retest reliability in the abstract.

I would see factorial invariance across gender or presence complication and Rasch model for this scale.

We agree with the reviewer that this would be of interest. However, there are some methodological difficulties that limit the possibility to do this. The first step in testing for factorial invariance across groups is to make separate CFA-models for each group and investigate model fit for each model. If the fit is poor in one or both models further testing for factorial invariance is not recommended (Brown, 2015, chapter 7). We estimated separate models for men and women and found poor fit in men and only moderate fit in women, probably partly caused by small sample size in each group. With less than 80 in each group we do not really have a large enough sample size to make separate models for men and women. A multiple-group CFA with sex specified as a grouping variable and all parameters allowed to vary freely (Test for equal form) also resulted in relatively poor fit. Since all consecutive models involved in further testing are based on comparison with the equal form model we do not think that results from analyses of measurement invariance across genders should be included in the paper.

Only 32 patients (22%) had complications and we do thus not have large enough sample size to test for invariance between patients with and without complications.

Rasch models are the simplest form of Item Response Theory (IRT)-models and can only be estimated when items are dichotomous. In this study the five items included in PAID-5 are ordinal with five categories and a Rasch model can thus not be estimated. There are however other types of IRT-models derived from the Rasch model for items with more than two categories, e.g. partial credit models and graded response models. Category characteristic curves and test information curve for a partial credit model are given below. For all items the probability of answering in higher categories increases with increasing levels of the underlying measured trait theta (diabetes related distress). We do not think that the IRT-analyses adds much information to the paper and suggest not to include it.

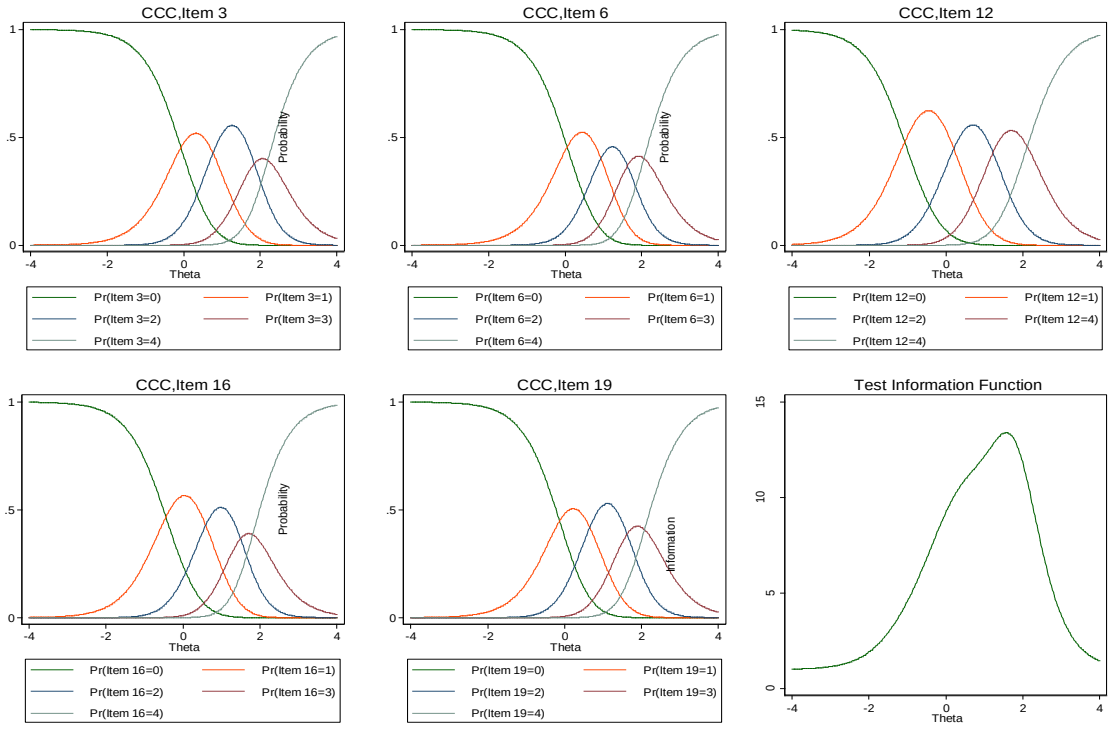

Figure: Category characteristics curves (CCC) and test information curve from partial credit model of the five items in PAID-5. Each CCC displays probability of answering in a specific category for the item as a function of theta, which is the level of the latent construct measured (diabetes distress).

It would be great if the authors calculate average variance extracted and composite reliability.

Average variance extracted (AVE) and composite reliability have now been added to Table 2.

I am not really sure how the authors want to generalize their consecutive sample size into the national. I would omit or revise the last sentence of conclusion

We agree that this was unclear and we have revised the sentence. This short questionnaire is useful in the national diabetes registry or population cohort studies as it enables increased knowledge on the prevalence of diabetes-related distress and thus inform guidelines and future interventions.

Introduction: I don't agree to separate introduction with background. It would be great to add some sentences on why a measure should adopt from a culture into another culture.

As suggested by both reviewers we have combined the introduction and background section and only kept the most relevant content for this validation study. In addition, we have added some sentences on cultural adaptation of patient reported outcome measures (page 4-5).

Method: the study is methodological not really test retest! Sample size justification is not really true from the Wolf's study. They said that a range of sample size requirements (i.e., from 30 to 460 cases), meaningful patterns of association between parameters and sample size. Typically, 5 to 20 respondents/parameter estimated per each parameters (Bentler, 1995).

We agree that it is misleading to state that we used a test-retest design since only a small fraction of the study population answered the questionnaire twice. We have now replaced the phrasing "test-retest" with "cross-sectional" in the first sentence in the Methods section (line 1/ para 1/ page 6).

When it comes to sample size, both the paper by Wolf et al and Chapter 10 in Brown, T.A (2015) states that we should not rely upon rules of thumb like 5 to 20 per parameter when we decide the sample size. According to both Wolf and Brown sample size calculations should be done for the particular model in question, either by using the Satorra-Saris method or by a Monte-Carlo simulation procedure suggested by Muthen and Muthen (2002). Since sample size calculation should always be done prior to data collection and actual testing of the model we have to do assumptions about the size of the factor loadings. The paper by Wolf et al has (in addition to other models) used the Monte-Carlo simulation method suggested by Muthen and Muthen and have calculated necessary sample size for one-factor models with 3-4, 6 and 8 items and loadings ranging from 0.50 to 0.80. These model criteria fit well with our 1-factor model for PAID5 with five items and relatively strong loadings. A one-factor model with 4 items and loadings equal to 0.5 requires a sample size of 190, while a model with six items and the same loadings requires a sample size of 90. We do thus think that the results from Wolf et al are adequate as arguments that the sample size in our study was large enough for a 1-factor CFA. We have changed the text in the methods-section to make this clearer (lines 3-9/ para 1/ page 6).

I cannot understand what is mean for Patients and or public!

We thought this was a mandatory section pointed out by the manuscript guidelines. However, we have removed the section as suggested.

Method of estimation for the CFA should be added. How the authors dealt with missing data in the CFA?

The method of estimation (maximum likelihood) has now been added to the methods section (Line 1-2/ para 1/Page 8).

In CFA-models missing values were handled with listwise deletion. This information is given on page 8, line 3-4 . Given that only two persons had missing values on at least one item we do not consider missing values to be an important issue in this study.

Results: please provide all factor loadings for the CFA model.

Standardized factor loadings have now been added to Table 2.

Table 1 should be omitted.

We have omitted Table 1 but would suggest to keep it as a supplementary table (Table S1).

Table 4 could be reported in the text.

Table 4 is now omitted and results are reported in the text (lines 258/ para 4/ page 9).

The authors should discuss why they added a covariance error terms between items 3 and 16. Brown (2015) highligts that you need a good justification to add correlated errors between some indicators of your construct and these correlation, that you should not only model them to reach the common cut-offs for good model fit and to be consistent with the rule you apply. He presents several potential reasons why such correlated errors can occur, like shared method variance due to different wording compared to other indicators, or specific item content (even you might not had this assumption a prior, I think you can defend them, if such reason is plausible). I can really recommend his chapter: Brown, T. A. (2015). Confirmatory factor analysis for applied research. Guilford Publications.

Correlated errors between two items may be necessary if correlation between items is caused by some other constructs than the construct we are trying to measure, in our case diabetes distress. Item 3 (“Feeling scared when you think about living with diabetes?” and item 16 (“Feeling that diabetes is taking up too much of your mental and physical energy every day?”) could be correlated independent of diabetes distress among persons who are anxious in general. But this is also true for item 6 (“Feeling depressed when you think about living with diabetes?”) for which error covariance with item 3 was found to improve model fit in the study by Lee et al. In our data error covariance between item 3 and item 6 also improved model fit, but not as much as for item 3 and item 16.

Since we do not have any strong arguments for including covariance between the two error terms we chose to present models both with and without error covariance and leave it up to the reader to judge which model they prefer to trust the most.

VERSION 2 – REVIEW

REVIEWER	Bik-Wai Bilvick Tai Caritas Institute of Higher Education, Hong Kong, China
REVIEW RETURNED	10-Sep-2018

GENERAL COMMENTS	In Table 1, the continuous variables (e.g. age, HbA1c, diabetes duration) are still not expressed in mean +/- SD appropriately.

	Significant grammatical mistakes and punctuation errors are still present throughout the manuscript that can adversely impact reading comprehension. The authors are advised to take serious action for revision before this manuscript can be considered for publication.

REVIEWER	AHP Iran
REVIEW RETURNED	14-Sep-2018

GENERAL COMMENTS	The authors have addressed my previous concerns

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

Significant grammatical mistakes and punctuation errors are still present throughout the manuscript that can adversely impact reading comprehension

We have proofread the text and corrected any spelling and grammar errors. The text has been examined by a person with English as her mother tongue, as well by a person with long experience of reading/writing scientific texts in English. We hope that the language is satisfactory.

In Table 1, the continuous variables (e.g. age, HbA1c, diabetes duration) are still not expressed in mean +/- SD appropriately.

We have checked and corrected Table 1.

VERSION 3 – REVIEW

REVIEWER	Bik-Wai Bilvick Tai Caritas Institute of Higher Education Hong Kong
REVIEW RETURNED	09-Nov-2018

GENERAL COMMENTS	My comments were adequately addressed by authors. Journal may need to review the English language in this manuscript again before publication.