Using core sets of the international classification of functioning, disability and health (ICF) to measure disability in vestibular disorders: Study protocol

Eva Grill[a,b,*], Joseph M. Furman[c], Alia A. Alghwiri[d], Martin Müller[a,b] and Susan L. Whitney[e]

[a]*Institute for Medical Information Processing, Biometrics and Epidemiology, Ludwig-Maximilians-Universität München, Munich, Germany*
[b]*German Center for Vertigo and Balance Disorders, Ludwig-Maximilians-Universität München, Munich, Germany*
[c]*Department of Otolaryngology, University of Pittsburgh, Pittsburgh, PA, USA*
[d]*Department of Physical Therapy, The University of Jordan, Amman, Jordan*
[e]*Department of Physical Therapy, University of Pittsburgh, Pittsburgh, PA, USA*

Abstract. Symptom frequency and severity in vestibular disorders often do not correlate well with patients' restrictions of activities of daily living and limitations of participation. Due to the lack of appropriate patient reported outcome measures (PRO), the extent of limitations and restrictions is mostly unknown. The International Classification of Functioning, Disability and Health (ICF) is a conceptual framework and classification to evaluate all aspects of health and disability. An ICF-based measure, the Vestibular and Participation Measure (VAP), was recently proposed. Also, an ICF Core Set for vertigo, dizziness and balance disorders was developed to describe what aspects of functioning should be measured. This study protocol describes the development and cross-cultural validation of a new measure, the VAP-extended (VAP-e), based on VAP and ICF Core Set on three continents. To determine objectivity and cross-cultural validity of the VAP and to find potentially redundant items, Rasch models will be used. The VAP-e will be created by modifying or adding items from the Activities and Participation and Environmental Factors component of the ICF Core Set. Reliability, objectivity and responsiveness of the VAP-e will be tested.

Keywords: Vertigo, dizziness, quality of life, activities of daily living, social participation, ICF

1. Introduction

Vestibular disorders are disabling conditions that can have a major impact on independence, employability, activities of daily living and overall quality of life [7,14,19,28]. More than ten percent of adults of working age report some degree of disability related to dizziness [20]. In the majority of cases, vertigo is recurrent; about 80% of the affected persons report severe impairment [21]. Taking into account all relevant aspects of functioning from the patient's perspective is therefore essential to evaluate new treatments and to convey the prognosis of the disease. There are many clinical tests and measures to evaluate impairments on the level of signs and symptoms such as reduced vestibulocular reflex gain, nausea, imbal-

*Corresponding author: Prof. Dr. Eva Grill, Institute for Medical Information Processing, Biometrics and Epidemiology, Ludwig-Maximilians-Universität München, Marchioninistr. 17, D-81377 München, Germany. Tel.: +49 89 2180 78223; Fax: +49 89 2180 78230; E-mail: eva.grill@med.uni-muenchen.de.

ance and dizziness. However, symptom frequency and severity often do not correlate well with a patients' ability to perform daily activities or participate in daily life [29]. Also, environmental factors such as the construction of buildings or the characteristics of natural surroundings [5] creating disorienting visual clues [25] or the movement of large objects, disorienting lights or sounds [6,22] can contribute to the burden of disability caused by vertigo.

Although the disabling impact of vertigo is obvious, the full extent and specific nature of limitations and restrictions in activities and participation is largely unknown. In a review of the most widely used vestibular measures Alghwiri et al. could show that none of those measures are specialized in activities and participation or in quantifying the effect of environmental factors [1]. Another systematic review on studies in vestibular rehabilitation found that interventions were mainly centred on symptoms but hardly on activities and participation outcomes [23]. Thus, it remains a challenge to evaluate the true disabling impact of vestibular disorders on functioning, i.e. on the totality of body functions and structures, activities and participation in interaction with the environment as proposed by WHO's International Classification of Functioning, Disability and Health (ICF).

The ICF is a conceptual framework to evaluate all aspects of health and disability. Sections or single categories of the ICF can be used to create new measures. To address issues of feasibility regarding the number of categories to be assessed and the user perspective, which typically consists of a health condition and/or situation perspective, ICF Core Sets can be defined. ICF Core Sets are selections of categories that describe the most relevant and the most common aspects of functioning as well as the most relevant environmental factors for persons with specific health conditions or in specific settings.

With the goal of putting salient ICF categories into action under the roof of a common scale, the new Vestibular and Participation Measure (VAP) was recently proposed [2]. Based on expert consensus and an iterative Delphi process, 34 items were derived from the ICF and validated in an ambulatory care setting with excellent results on reliability, concurrent and convergent validity. Still, objectivity of the scale remains to be shown. Objectivity is proven if a scale will satisfy the axioms of additive conjoint measurement [18], i.e. if item responses of a scale can be added to a meaningful single score. Also, cross-cultural validity has not been examined. Finally, a questionnaire

with 34 items might be too comprehensive for clinical routine. The analysis of objectivity also yields information about redundant items that may be dropped without loss of information and may thus lead to a more parsimonious scale.

Although there is vast literature on the application of the ICF in various settings and health conditions [10, 12,13], the VAP is one of the few examples of its use in vestibular disorders. To fill in this existing gap, an ICF Core Set for vertigo, dizziness and balance disorders was developed in an international consensus process integrating information from empirical qualitative and quantitative studies [9]. This ICF Core Set contains 31 categories from the components Body Function/Body Structures, 40 categories from the component Activities and Participation and 29 categories from the component Environmental Factors. Also, a short form, the Brief ICF Core Set for Vertigo, was created. Experts felt that the selected categories from Environmental Factors were of specific relevance, either as facilitating factors (e.g., products and technologies such as walking aids) or as triggers or barriers. During the international approval process of the ICF Core Sets for Vertigo the participating experts recommended that the next step should be the development of a new PRO based on Core Set categories and on existing measures.

In the literature, several obstacles are brought forward when discussing the application of the ICF or ICF Core Sets in clinical practice [11]. One of them is the lack of standardized measures. The ICF provides so called qualifiers for scoring single categories. Qualifiers are codes ranging from 0 to 4 quantifying the extent of a problem in reference to the data of an external population. Still, the ICF categories cannot be applied as measures "as they are". They are lists of concepts indicating what could be measured but not how to measure them. How to select appropriate measurement instruments based on ICF Core Sets requires several decisions, e.g. on the level of detail (single item or item battery), on the addressee (self-report by patient or clinical assessment by health professional), or on the desired psychometric qualities (sensitivity for change, discrimination). New measures can be created by modifying or adding items to existing measures. New items can be written based on ICF categories. For cross-cultural adaptation of the ICF, revisions are already available in multiple languages.

Creation of a PRO is an iterative process with several steps going from the identification of concepts to item generation and instrument formatting to the assessment of the measurement properties of the instru-

ment [8]. Despite content validity, new items and new scales based on the ICF Core Set for vertigo still have to undergo testing to prove their reliability and validity clinically [9].

Regarding the application of the ICF for vestibular disorders, comparing and combining the existing version or a new, modified version of the VAP with the ICF Core Set might lead to a conceptually and psychometrically refined measure.

There are therefore three main objectives of this study: (1) to determine objectivity and cross-cultural validity of the VAP, (2) reduce the number of items in the VAP to create a short form more amenable to clinical use (the VAP-s), and (3) to modify and extend the VAP by modifying items, adding items from the Activities and Participation and Environmental Factors component of the ICF Core Set, and evaluating reliability, objectivity and responsiveness of the new measure called the VAP-extended (VAP-e).

2. Methods

The study will have two phases. The first phase serves to test for cross-cultural validity of the VAP, potentially also to reduce the number of items of the VAP. The second phase serves to extend the VAP by ICF categories from the ICF Core Set and test this extended version psychometrically. Figure 1 gives a detailed overview of the two phases and the steps of the project.

2.1. Phase 1. Testing objectivity and cross-cultural validity of the VAP

2.1.1. Study design and sample

The cross-cultural validation study will be a multicenter survey with three centres on three different continents, including patients from the United States (Pittsburgh), Germany (Munich) and Jordan (Amman).

Individuals with vestibular disorders between the age of 18–100 who have adequate command of the language of the respective country and provide informed consent will be included. Individuals will be excluded if they are unable to complete the questionnaire because of cognitive barriers or if an acute medical condition is associated with dizziness and requires immediate attention (e.g. acute myocardial infarction or stroke). Participants will be asked to complete the VAP during their initial visit to the University of Pittsburgh Medical Centers Balance and Vestibular Clinic, the
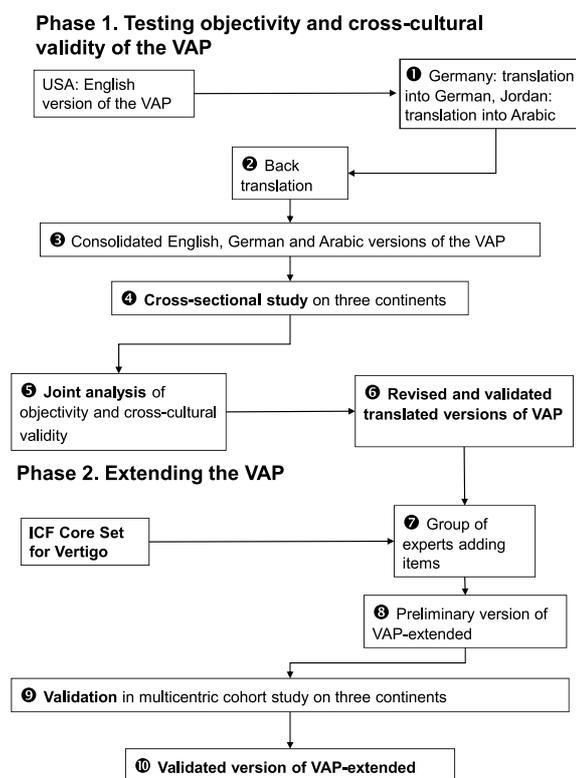

Fig. 1. Flow chart of the testing and extension of the Vestibular and Participation Measure (VAP).

outpatient dizziness clinic at the German Center for Vertigo and Balance Disorders (IFBLMU) at Ludwig-Maximilians-Universität in Munich, and the Middle East Center for Hearing and Balance, Amman. In each centre, patients will be asked to complete the questionnaires before diagnostic procedures start.

The study was given ethical approval by the Institutional Review Boards of the respective institutions.

2.1.2. Measures

The VAP is a 34-item self-report questionnaire based on the Activities and Participation component of the ICF. This measure evaluates the effect of dizziness and/or balance problems on the ability to perform activity and participation tasks. The difficulty is rated without the assistance of other persons on each task. The response scale of the VAP is set as a 5-point scale indicating the level of difficulty (none = 0, mild = 1, moderate = 2, severe = 3, unable to do = 4, and not applicable). The authors report that a total score of the VAP can be obtained by calculating the average of the item scale values after excluding the not applicable items. Reliability, concurrent and convergent valid-

ity of the VAP were confirmed in patients attending a tertiary balance clinic [2].

In addition, convergent validity will be appraised by the Dizziness Handicap Inventory (DHI) [15]. The DHI addresses self-perceived handicap due to vestibular disorders. The score ranges from 0 to 100, with a higher score indicating greater handicap.

To describe the study sample, vestibular test results symptom severity as assessed by the Vertigo Symptom Scale VSS [29], the diagnosis made by the neurotologist based on common diagnostic guidelines [4], and anamnestic (medication, comorbidities) and demographic data (gender, age, education) will be recorded.

2.1.3. Statistical analyses

Data from individual countries will be analysed separately and as pooled data. Analyses make use of the theory of the Rasch unidimensional measurement model [24]. This theory explains how items on a scale perform and how the tested persons' ability relates to the properties of the scale. The Rasch model allows validating a scale with dichotomous or polytomous items in a way that it is possible to sum up single item values to a summary score, therefore, to calculate change scores and effect sizes in a statistically meaningful way. The basic assumption of the Rasch model is that it is more likely to pass an easy item as compared to a difficult item, and that a person with high ability is more likely to pass an item than a person with low ability. The probability of passing an item is defined as a logistic function of the difference between person ability and item difficulty.

If the observed responses of an item set fit the Rasch model, that is, if easy items are passed by more patients and more able persons pass more items, then the item set is unidimensional and constitutes a Rasch scale. This can be verified by specific goodness-of-fit statistics. It can then be assumed that the items of the scale measure one single trait. This is confirmed by analysis of local independence of items. If the underlying single trait (the so called Rasch factor) is removed, there should be no residual associations left. Additionally, another important property to be examined is if a scale functions consistently across different sub-groups of patients, that is, if an item is equally easy or difficult for men and women, for younger and older persons. Thus each item is examined for differential item functioning (DIF) across levels of person factors.

The set of items will be fitted to the Rasch model. Goodness of fit will be examined by overall fit statistics and individual item-fit statistics. We will use three over-all fit statistics; the two item-person interaction statistics (for item and for person fit) are z-distributed, that is, a mean of zero and a standard deviation of 1 indicate perfect fit to the model, the item-trait interaction statistic has a chi-square distribution. Good fit would be indicated by a non-significant value of the item-trait interaction statistic, indicating that the rank of the items is constant irrespective of the level of the underlying trait. The individual item-fit statistics indicate how well an item fits into the scale. We will report fit residuals as the total of individual person and item deviations from the model that should lie within the range of ± 2.5 and the chi-square statistic, which should be non-significant after adjustment for multiple testing.

To demonstrate if the scales function independently from person factors, we will examine if at any given level of person ability the probability to pass an item is equal across strata of a personal factor. For example, when examining the category walking (the item), a woman with a certain level of limitation in mobility (the person ability) will have the same probability to be able to walk (to pass the item) as compared to a man with this level of limitation, if the Rasch assumptions hold.

Thus, each item is examined for DIF across strata. The person-response residuals for each item show how much a person at a certain ability level departs from the response expected at this level of ability. Person factor and class intervals are treated as factors of an ANOVA. Their ANOVA statistics should be non-significant after appropriate adjustment for multiple testing. If there is uniform DIF of some of the items, that is, the detected difference in response across strata is the same regardless of the level of ability; those items can be adjusted to vary by person factor. In contrast, nonuniform DIF, that is, varying differences in response across strata, cannot be accounted for. In this study, we will examine DIF for country by dividing the sample into patients from the US, Germany or Jordan, for age by dividing the sample into two age groups, and for gender.

As a final test of unidimensionality, person-item deviation residuals will be examined by principal components analysis [27]. A person separation index can be used to show how patients have been spread out along the scale defined by the items, thus indicating how well the scale differentiates between patients. The person separation index is close to 0 when all the persons are in a similar location and approaches 1 the more they spread across the item continuum.

To create the short form of the VAP (VAP-s), items will be examined according to their location on the

item continuum. An item might be a candidate for deletion if its location equals the location of another item. An item might also be deleted from a scale if it does not fit well onto the scale. In case of several misfitting items, creation of subscales can be the solution.

Significance level will be set at 0.05. The Bonferroni method will be applied as adjustment for multiple testing, yielding a significance level of 0.05/k where k is the number of tests carried out simultaneously. Missings in single VAP items will be resolved by multiple imputation. Sensitivity analyses will be carried out with imputed and not-imputed data sets. Rasch analysis will be carried out with the RUMM2030 package (Perth, RUMM Laboratory), bivariate statistical analyses will be carried out with SAS V9.3 (Cary, NC).

2.1.4. Sample size

In accordance with guidelines for Rasch models [17] 150 participants per country are sufficient to provide stable estimates.

2.2. Phase 2. Extending the VAP

Item writing

The ICF Core Set for vertigo contains a list of 40 categories for the component Activities and Participation (A&P) and a list of 29 categories for the component Environmental Factors (EF). Activity and Participation categories will be compared with the modified VAP and items will be added if necessary. Original items of the VAP can be left as they were, but they can also be deleted or modified.

A group of experts will generate the wording of items based on the pool of additional A&P and EF categories from the ICF Core Set. Comments retrieved from qualitative interviews with patients with vestibular disease [19] will be used for additional guidance. Wording will be developed in English, German and Arabic. The recommended cross-cultural adaptation process [3] will be followed, namely translation from the English language into German and Arabic, synthesis of translations, blinded back translation into English, evaluation of translations by an expert panel deciding on equivalence with re-translation and correction if necessary.

In order to review the wording, patients at each site will be asked to proof the items. They will be asked to evaluate the items according to relevance, applicability, comprehensibility, and appropriateness.

The new items will then be added to the modified version of the VAP to form the VAP-extended (VAP-e).

The Brief ICF Core Set for Vertigo comprises 10 categories from the component A&P, and six categories from the component EF. For a short form of the VAP-e, items of the VAP-e will be deleted if they are not corresponding to categories contained in the Brief Core Set.

2.3. Validation of the VAP-extended

2.3.1. Study design and sample

Procedures will be carried out in analogue to Phase 1. Data will be collected in a multicentre cohort study with the three centres on three different continents, including patients from the United States (Pittsburgh), Germany (Munich) and Jordan (Amman), including individuals with vestibular disorders between the age of 18–100 who have adequate command of the language of the respective country and give informed consent. The study protocol will be sent to the review boards of the respective institutions for approval.

2.3.2. Data collection procedures

To evaluate responsiveness and sensitivity to change, the measures will be completed at the patients' first visit to the clinic, and again six months after via postal survey. Non-responders will be contacted by phone for motivation and to get non-responder information (reason for non response, e.g. vital status or new address). To evaluate test-retest reliability, a sample of patients will be asked to complete the VAP-e twice on the same day. The order of items will be changed to avoid recall bias.

2.3.3. Measures

To describe an overall impression of a participant's health, participants will be asked to appraise their health on a 5-point-Likert scale ranging from one to five where one indicates best health. In addition, convergent validity will be appraised by the Dizziness Handicap Inventory (DHI) [15]. The DHI addresses self-perceived handicap due to vestibular disorders. The score ranges from 0 to 100, with a higher score indicating greater handicap. In analogue to Phase 1, symptom severity will be assessed by the Vertigo Symptom Scale VSS [29]. Likewise, the diagnosis made by the neurotologist based on common diagnostic guidelines [4], and anamnestic (medication, comorbidities) and demographic data (gender, age, education) will be recorded.

2.3.4. Statistical analysis

The intraclass correlation coefficient (ICC) and its 95% confidence interval will be used to estimate test-

retest reliability of the VAP-e total score. An ICC of greater than 0.75 indicates excellent reliability, 0.40 to 0.74 indicates fair to good reliability, and less than 0.40 indicates poor reliability [16]. The Cohen's kappa statistics (weighted and unweighted) will be used to estimate the agreement (above chance level) between person ratings on successive test administrations for individual items. Kappa ranges from 0 to 1, in which 0 indicates no agreement and 1 indicates perfect agreement. A kappa of 0.75 indicates excellent agreement, 0.40 to 0.74 indicates good agreement, and less than 0.40 indicates poor agreement [16]. General linear models with random effect coefficients for country effects will be used to evaluate responsiveness and sensitivity to change.

To examine if items of the VAP-e and its short form can be added to a meaningful single score, Rasch analyses will be conducted accordingly, as described above, for both the VAP-e and its short form. Since we hypothesize activity and participation and environmental factors forming two distinct scales, models will be carried out separately for both components.

2.3.5. Sample size

In an ambulatory setting, reliability testing with more than two replications seems hardly feasible. Also, it has been shown that in the case of adequate reliability (ICC $\geqslant 0.6$), more than two replications do not add substantially to the test efficiency. Assuming an ICC coefficient of at least 0.6, sample sizes of 50 patients tested twice (i.e. 100 observations) per centre will be sufficient to minimize the variance of the coefficient (< 0.008 for an ICC of 0.6, < 0.001 for an ICC of 0.9) [26].

Rasch analysis will be carried out with the RUMM 2030 package (Perth, RUMM Laboratory). SAS V9.3 (Cary, NC) will be used for all validation analyses.

3. Conclusion

Vestibular disease can have a dramatic impact on quality of life and functioning of an individual, regardless of objective clinical tests and symptoms. The ICF is a common framework for all health professionals for the coding and assessment of functioning. The Vestibular Activities and Participation Measure (VAP) was the first to use the ICF to introduce activities and participation outcomes to vestibular research. Psychometrical testing of the VAP and its extension along the philosophy of the ICF might result in valid and reliable measures that can be used to assess the success of therapy options in vestibular disease.

Acknowledgments

References

[1] A.A. Alghwiri, G.F. Marchetti and S.L. Whitney, Content comparison of self-report measures used in vestibular rehabilitation based on the international classification of functioning, disability and health, *Phys Ther* **91** (2011), 346–357.

[2] A.A. Alghwiri, S.L. Whitney, C.E. Baker, P.J. Sparto, G.F. Marchetti, J.C. Rogers and J.M. Furman, The development and validation of the vestibular activities and participation measure, *Arch Phys Med Rehabil* **93** (2012), 1822–1831.

[3] D. Beaton, C. Bombardier, F. Guillemin and M. Ferraz, Guidelines for the process of cross-cultural adaptation of self-report measures, *Spine* **25** (2000), 3186–3191.

[4] A. Bisdorff, M. Von Brevern, T. Lempert and D.E. Newman-Toker, Classification of vestibular symptoms: towards an international classification of vestibular disorders, *J Vestib Res* **19** (2009), 1–13.

[5] T. Brandt, F. Arnold, W. Bles and T.S. Kapteyn, The mechanism of physiological height vertigo. I. Theoretical approach and psychophysics, *Acta Otolaryngol* **89** (1980), 513–523.

[6] A.M. Bronstein, Vision and vertigo: Some visual aspects of vestibular disorders, *J Neurol* **251** (2004), 381–387.

[7] A.M. Bronstein, J.F. Golding, M.A. Gresty, M. Mandala, D. Nuti, A. Shetye and Y. Silove, The social impact of dizziness in London and Siena, *J Neurol* **257** (2010), 183–190.

[8] FDA, Guidance for industry: patient-reported outcome measures: Use in medical product development to support labeling claims: draft guidance, *Health Qual Life Outcomes* **4** (2006), 79.

[9] E. Grill, A. Bronstein, J. Furman, D.S. Zee and M. Muller, International Classification of Functioning, Disability and Health (ICF) Core Set for patients with vertigo, dizziness and balance disorders, *J Vestib Res* **22** (2012), 261–271.

[10] E. Grill, B. Lipp, C. Boldt, G. Stucki and E. Koenig, Identification of relevant ICF categories by patients with neurological conditions in early post-acute rehabilitation facilities, *Disabil Rehabil* **27** (2005), 459–465.

[11] E. Grill and M. Muller, The ICF and Florence Nightingale – bringing data to statistical proof, *J Rehabil Med* **43** (2011), 1041–1042.

[12] E. Grill, M. Quittan, V. Fialka-Moser, M. Muller, R. Strobl, N. Kostanjsek and G. Stucki, Brief ICF Core Sets for the acute hospital, *J Rehabil Med* **43** (2011), 123–130.

[13] E. Grill, R. Strobl, M. Muller, M. Quittan, N. Kostanjsek and G. Stucki, ICF Core Sets for early post-acute rehabilitation facilities, *J Rehabil Med* **43** (2011), 131–138.

[14] N. Guinand, F. Boselie, J.P. Guyot and H. Kingma, Quality of life of patients with bilateral vestibulopathy, *Ann Otol Rhinol Laryngol* **121** (2012), 471–477.

[15] G.P. Jacobson and C.W. Newman, The development of the Dizziness Handicap Inventory, *Arch Otolaryngol Head Neck Surg* **116** (1990), 424–427.

[16] J. Landis and G. Koch, The measurement of observer agreement for categorical data, *Biometrics* **33** (1977), 159–174.

[17] J. Linacre, Sample Size and Item Calibration Stability, *Rasch Measurement Transactions* **7** (1994), 328–331.

[18] R.D. Luce and J.W. Tukey, Simultaneous conjoint measurement: A new scale type of fundamental measurement, *Journal of Mathematical Psychology* **1** (1964), 1–27.

[19] M. Mueller, E. Schuster, R. Strobl and E. Grill, Identification of aspects of functioning, disability and health relevant to patients experiencing vertigo: A qualitative study using the international classification of functioning, disability and health, *Health Qual Life Outcomes* **10** (2012), 75.

[20] I. Nazareth and L. Yardley, A clinical approach to dizziness, *J Clin Outcomes Med* **9** (2002), 159–167.

[21] H.K. Neuhauser, M. von Brevern, A. Radtke, F. Lezius, M. Feldmann, T. Ziese and T. Lempert, Epidemiology of vestibular vertigo: A neurotologic survey of the general population, *Neurology* **65** (2005), 898–904.

[22] M. Pavlou, The use of optokinetic stimulation in vestibular rehabilitation, *J Neurol Phys Ther* **34** (2010), 105–110.

[23] F. Porciuncula, C.C. Johnson and L.B. Glickman, The effect of vestibular rehabilitation on adults with bilateral vestibular hypofunction: A systematic review, *J Vestib Res* **22** (2012), 283–298.

[24] G. Rasch, Probabilistic models for some intelligence and attainment tests, University of Chicago Press, Chicago, IL, 1980.

[25] M.S. Redfern, L. Yardley and A.M. Bronstein, Visual influences on balance, *J Anxiety Disord* **15** (2001), 81–94.

[26] M.M. Shoukri, M.H. Asyali and A. Donner, Sample size requirements for the design of reliability studies: Review and new results, *Stat Methods Med Res* **13** (2004), 251–271.

[27] E.J. Smith, Detecting and evaluating the impact of multidimension-ality using item fit statistics and principal component analysis of residuals, *J Appl Meas* **3** (2002), 205–231.

[28] J. White, P. Savvides, N. Cherian and J. Oas, Canalith repositioning for benign paroxysmal positional vertigo, *Otol Neurotol* **26** (2005), 704–710.

[29] L. Yardley, E. Masson, C. Verschuur, N. Haacke and L. Luxon, Symptoms, anxiety and handicap in dizzy patients: development of the vertigo symptom scale, *J Psychosom Res* **36** (1992), 731–741.