Title: A critical review of the psychometric properties of the Nijmegen Questionnaire for hyperventilation syndrome.

Authors: Vickie Li Ogilvie (NZRP, BHSc (Physiotherapy), PGDip Rehabilitation). Respiratory Physiotherapist, Acute Allied Health, Middlemore Hospital, Auckland.

Paula Kersten (PhD, PGCert Academic Practice, MSc, BSc (Physiotherapy)). Professor, Centre for Person Centred Research, School of Rehabilitation & Occupation Studies, Auckland University of Technology, Auckland.

The corresponding author and contact details: Vickie Li Ogilvie

Acute Allied Health, Sir Edmund Hillary Building, Middlemore Hospital, 100 Hospital Road, Papatoetoe 2025, New Zealand.

vickie.li.ogilvie@middlemore.co.nz

+64 9 2760210

Source of funding:

Partial funding was granted for this paper from the Cardiothoracic Special Interest Group, Physiothearpy New Zealand.

Abstract

The Nijmegen Questionnaire is commonly used by physiotherapists and other health professionals in the clinical and research settings. This outcome measure was developed by a group of researchers at the Nijmegen University in the Netherlands as a screening tool for the hyperventilation syndrome in the 1980s. However, the literature that supports the efficacy of its use is scarce. This paper examines the evidence in relation to the conceptual basis, validity, and reliability of the Nijmegen Questionnaire. A systematic review of the literature is carried out to identify studies that are related to the above measurement properties for the questionnaire. Studies identified are evaluated for their methodological qualities using the COSMIN checklist. The clinical utility of this instrument is also discussed. Issues associated with the development and validating process of this outcome measure are identified. There is also a lack of evidence in cultural validation given that the Nijmegen Questionnaire is developed in the Netherlands. While this is the only questionnaire currently available that is designed specifically for the screening of hyperventilation syndrome, administrators need to be aware of the issues identified in relation to validity and reliability when interpreting the results. Applying more robust validating processes to establish the efficacy of the Nijmegen Questionnaire appears to be a priority for researchers in order to improve the quality of health services for individuals suffering from hyperventilation syndrome.

Key words

Nijmegen questionnaire

Self evaluation of breathing questionnaire

Rowley breathing self efficacy scale

Breathing pattern disorders

Dysfunctional breathing

Hyperventilation

Questionnaire

Outcome measures

Assessment

Reliability

Validity

Introduction

Hyperventilation syndrome (HVS) is a breathing pattern disorder which is often undiagnosed due to its multi-systemic and apparently unrelated symptoms (Mooney and Candy 2008, van Doorn et al 1983). HVS sufferers are regarded as high healthcare users due to the involvement of various medical or surgical services and array of investigations (Chaitow et al 2002, Lum 1975). Mooney and Candy (2008) have demonstrated that the financial implications are significant for both the patients with HVS and their healthcare providers.

Early diagnosis and implementation of individualised physiotherapy education and treatment are proposed as cost effective management approaches for patients with HVS (Mooney and Candy 2008). Diagnostic and screening tools for HVS include the hyperventilation provocation test (HVPT) and formulated questionnaires (Vansteenkiste et al 1991). HVPT is criterion for diagnosis and requires an individual to hyperventilate for few minutes to reproduce presenting symptoms of HVS (Hornsveld et al 1996). Outcome measures that assess hyperventilation and dysfunctional breathing include the Nijmegen Questionnaire, 33-item Hyperventilation Questionnaire (HVQ), and the Self Evaluation of Breathing Questionnaire (SEBQ) (Rapee and Medoro 1994, Courtney and Greenwood 2009, Vansteenkiste et al 1991). However, only the Nijmegen Questionnaire is suggested in the literature to be suitable for screening of HVS in adults (van Dixhoorn and Duivenvoorden 1985). Another questionnaire, the Rowley Breathing Self-Efficacy scale (RoBE scale) (Rowley and Nicholls 2006) is associated with the assessment of people with breathing pattern disorders but, its focus is on investigating the individual's ability to control their symptoms in relation to breathing pattern disorders. This leaves the

Nijmegen Questionnaire, which is widely used for the detection and diagnosis of HVS (van Dixhoorn and Duivenvoorden 1985).

The Nijmegen Questionnaire (see Appendix) is a short, self-administered patient reported outcome measure consisting 16 HVS related complaints. The frequency of occurrence can be rated on a five-point ordinal scale (0: never, 4: very often) (van Dixhoorn and Duivenvoorden 1985, van Doorn et al 1982). A score above 23/64 is a positive screening of HVS (Garssen et al 1984, van Doorn et al 1983, Vansteenkiste et al 1991). This questionnaire is non-invasive in nature compared to the HVPT. It is considered to be an accurate indicator for hyperventilation within the multidisciplinary setting (Chaitow et al 2002). Routine application of this tool is common in New Zealand physiotherapy practice of patients with breathing pattern disorders including HVS. However, data on the validity and reliability of the tool have not been synthesised to date.

In this paper, we report findings from a systematic review of the evidence for the validity and reliability of the Nijmegen Questionnaire. The conceptual basis of the Nijmegen Questionnaire is also explored using the criteria compiled by the Scientific Advisory Committee of the Medical Outcomes Trust (2002). The mechanism and difficulties surrounding the integration of this outcome measure in relation to its clinical utility within the physiotherapy outpatient setting are also explored at the end of the Results section.

Before moving into the Method section, a brief definition of all measurement properties relating to our evaluation are outlined in the following paragraphs for the purpose of this review.

Validity

The examination of validity is paramount in the process of test development and it involves a number of sequential steps before the final goal of creating a valid outcome measure is achieved (Laver Fawcett 2007, Pallant 2001). The basic definition of validity in the subject field of outcome measurement is the degree to which a scale is measuring what it is designed to measure (Hambleton and Jones 1993, McDowell 2006, Streiner and Norman 2008). Streiner and Norman (2008) further define the process of validating a test as a means to establish the level of confidence we can assume when inferences are made about individuals based on their scores from that outcome measure. Validity can be grouped into three types (see Table 1): content, construct, and criterion validity, with the latter looking at specificity and sensitivity specifically (Bowling 1997, McDowell 2006, Pallant 2001, Streiner and Norman 2008).

Content validity

In the literature, it is suggested that the content validity of a scale relates to whether the items or questions included are representative of all the attributes to be evaluated within the specified conceptual basis while meeting the objectives identified for the given instrument (Bowling 1997, McDowell 2006). Additionally, Streiner and Norman (2008) suggest the inclusion of a representative sample in the process of test development can lead to more accurate inferences of individuals being evaluated that are applicable to variety of circumstances, hence increasing the content validity of the instrument developed.

A sound conceptual basis is essential in the development of a health related outcome measure (McDowell 2006). The various aspects of a specified conceptual model articulate the concepts and populations that a measuring tool intends to

evaluate and the relationships between the concepts (Scientific Advisory Committee of the Medical Outcomes Trust 2002). McDowell (2006) explains that a defined conceptual basis of a measure supports its content and allows the results obtained to be interpreted alongside a broader body of theory that is associated with the conceptual definition.

Construct validity

The presence of HVS is recognised through the identification of a variety of physical and psychological symptoms (Grossman and de Swart 1984). Such constellations of symptoms of HVS are considered by Streiner and Norman (2008) as hypothetical constructs. The process of construct validation of an outcome measure is complex because there is no one single test or criterion standard to follow (McDowell 2006). Construct validity of an instrument can only be established through an on-going process of learning, understanding, and testing of the constructs (McDowell 2006, Streiner and Norman 2008). Test developers need to look for a cumulative pattern of evidence to ascertain whether the emerging outcome measure relates to the theoretical constructs proposed when assessing the construct validity (Laver Fawcett 2007).

Criterion validity

Criterion validity is defined traditionally as the correlation of an instrument with another measuring tool that is considered the ‘gold standard’ in the same field (Bowling 1997, McDowell 2006, Streiner and Norman 2008). The comparison could be used formatively when developing a new tool to guide the items selection process by recognising the elements that correlate optimally with the criterion/’gold standard’ (McDowell 2006). When assessing concurrent validity (a form of criterion validity), the

researchers correlate a new measure with a measure that has been validated, i.e. both measures are administrated concurrently (Streiner and Norman 2008).

Cultural validity

The cultural background of the individual being evaluated can affect test administration and data interpretation (Laver Fawcett 2007). Health professionals need to select a valid and reliable assessment tool that is also culturally relevant to the people being assessed (Høegh and Høegh 2009). There are existing cross-cultural adaptation guidelines and processes in the literature that can help enhance the level of cultural validity or adaptability of a measurement tool (Beaton et al 2000, Høegh and Høegh 2009). Cultural validation process is not simply having the outcome measure translated to a different language; it is also to ensure the conceptual foundation of the outcome remains unchanged after the necessary adaptation of individual items (Beaton et al 2000).

Reliability

The various types of reliability in relation to patient reported outcome measure are internal consistency and test-retest reliability (Bowling 2001). Intenal reliability is the degree of the interrelatedness among the items, whereas test-retest reliablity is the extent to which scores on the same version of questionnaire for people who have not changed are the same for repeated measurement over time (Mokkink et al. 2010).

Table 1. Definitions of different measurement properties and various aspects of the validity domain

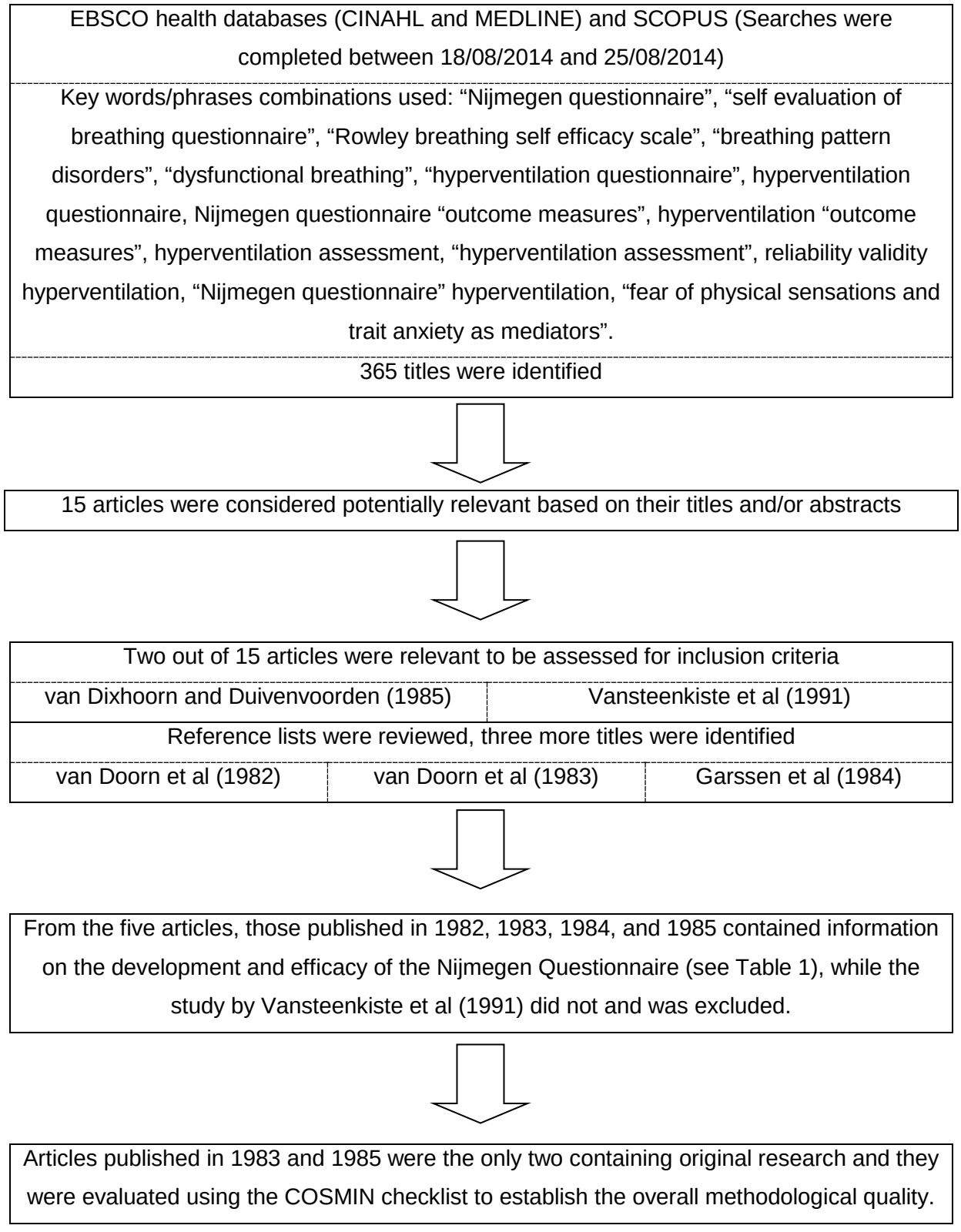
Domain	Measurement Property	Aspect	Definition
Reliability		Test-rest reliability	The degree to which the measurement is free from error and scores recorded have not changed are the same for repeated measurement over time.
Validity	Content validity	___	The degree to which the content of an instrument is an adequate reflection of the construct to be measured
	Criterion validity	___	The degree to which the scores of an instrument are an adequate reflection of a 'gold standard'
	Construct validity	Hypotheses testing	The degree to which the scores of an instrument are consistent with hypotheses with regard to differences between groups
		Structural validity	The degree to which the scores of an instrument are an adequate reflection of the dimensionality of the construct to be measured

Note. Only the measurement properties that are included in the two studies are presented here. Adapted from *Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist* by CB Terwee, LB Mokkink, DL Knol, R Ostelo, LM Boutex, and H de Vet (2012).

Methods

A literature search of the electronic databases (EBSCO Health databases, including CINAHL and MEDLINE) and health related citation index (SCOPUS) was undertaken to identify all articles that examined the validity and reliability of the Nijmegen Questionnaire for hyperventilation syndrome in adults, in addition to articles that were relevant to the development of the tool. Specific key words/phrases combinations were used for the electronic searches (see Figure 1). There was no limitation set on publication date. Papers published up till 25th August 2014 were included. The titles and abstracts of each paper form the initial searches except duplicates were reviewed for relevance. The full text was read if information provided in the abstract was insufficient. The reference lists of the articles identified from the initial searches were hand-searched to identify potential relevant titles. Studies were included if: (1) the aim of the study was to examine the *psychometric properties* (e.g. validity, reliability, sensitivity, or responsiveness) of the Nijmegen Questionnaire for hyperventilation syndrome in adults; (2) the study contained information relevant to the *development* of the Nijmegen Questionnaire for hyperventilation syndrome in adults. Studies were excluded if: (a) the study was puplished in languages other than English or Dutch (although there were none); (b) the participants of the study were younger than 18 years of age; (c) the participatns of the study were diagnosed with any organic cardiac, neruological, or respiratory disease.

Figure 1. Flow diagram showing the selection process of articles for A critical review of the Nijmegen Questionnaire in relation to hyperventilation syndrome.



Critical evaluation of the studies that met our review criteria was guided by the COSMIN checklist (Consensus-based Standards for the selection of health status Measurement INstruments), a standardised tool recommended for evaluating the methodological quality of studies concerning measurement properties (Mokkink 2010, Terwee et al 2012).

Results

An overview of the paper selection process is shown in Figure 1. A total of 365 articles were generated electronically after discarding duplicates. Fifteen were identified as potentially relevant to this review based on their study titles and/or abstracts. Thirteen of these were rejected based on our exclusion criteria. The two remaining articles were read in their entirety and reference list checking led the researchers to three more titles. Upon further inspections, four of the five articles provided information about the development of the Nijmegen Questionnaire and its validity and reliability data (see Table 2 for a summary of studies included in this review) of the tool. Translation of Dutch papers was provided by one of the authors of this paper, whose first language is Dutch. Only two of the four articles contained original research. These two research studies were led by van Doorn (1983) and van Dixhoorn (1985) respectively. A critical evaluation of these two studies was guided by the COSMIN checklist (see Table 3 for a summary of the evaluation).

Table 2: Summary of studies in relation to the critical review of the Nijmegen Questionnaire

Authors	Year	Study title	Purpose of the study	Results
van Doorn, Folgering, and Colla.	1982	Control of the end-tidal PCO2 in the hyperventilation syndrome: Effects of biofeedback and breathing instructions compared	To evaluate the efficacy of a behavioural management of HVS	Behavioural management supplemented with explanations about the mechanisms of HVS and coping strategies are useful.
van Doorn, Colla, and Folgering.	1983	Een vragenlijst voor hyperventilatieklachten [A questionnaire for hyperventilation symptoms]	To investigate if a short questionnaire in which patients are asked to report the frequency of 16 common hyperventilation symptoms is useful	The questionnaire is useful in patient screening and the provocation test can be used to rule out false positives.
Garssen, Colla, van Dixhoorn, van Doorn, Folgering, Stoop, and de Swart.	1984	Het herkennen van het hyperventilatiesyndroom [Recognising the hyperventilation syndrome]	To assess and review the NQ	*The NQ is able to discriminate (23 as the cut-off score) between individuals with and without HVS.
van Dixhoorm, and Duivenvoorden	1985	Efficacy of Nijmegen Questionnaire in recognition of the hyperventilation syndrome	To establish the differentiating ability of the NQ by comparing individuals with and without HVS	The NQ is a suitable screening tool for early detection of HVS and an aid in diagnosis and therapy planning.

Note. HVS = hyperventilation syndrome; NQ = Nijmegen Questionnaire. *This study result was adapted from the study by van Doorn and colleague (1983).

Table 3: Summary of study evaluation using the COSMIN checklist in relation to the Nijmegen Questionnaire

Evaluated measurement properties	Studies with original research		Overall quality scores	Questions for each property													
	Van Doorn, Colla, Folgering (1983)	Van Dixhoorn, Duivenvoorden (1985)		1	2	3	4	5	6	7	8	9	10	11	12	13	14
Reliability																	
	√		Poor	Good	Fair	Excellent	Poor	Excellent	Excellent	Good	Excellent	Good	Excellent	Poor	Poor	Poor	Excellent
Content validity																	
	√		Poor	Fair	Poor	Good	Fair	Poor									
Structural validity																	
		√	Poor		Good	Fair	Poor	Excellent	Excellent	Poor							
Hypotheses testing																	
		√	Fair	Good	Fair	Excellent	Fair	Good	Excellent	N/A	N/A	N/A	Excellent				
Criterion validity																	
	√		Fair	Good	Fair	Excellent	Excellent	Excellent	N/A	Excellent							

Note. Only the measurement properties that are included in the two studies are presented here. Excluded properties are internal consistency, measurement error, cross-cultural validity, and responsiveness. √ denotes the study that tested the specified measurement property. Each property has different number of questions within the COSMIN checklist as shown in the table. N/A indicates a lack of information from the study to answer the question listed. Adapted from *Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist* by CB Terwee, LB Mokkink, DL Knol, R Ostelo, LM Boutex, and H de Vet (2012).

Content validity

The conceptual and empirical basis for the inclusion of the 16 items was published over three decades ago (van Doorn et al 1982). The researchers stated that the items were chosen out of a list of 45 complaints that were regarded as associated with HVS for their clinical relevance by a group of specialists from various disciplines. These items were tested in two other studies with 40 and over 200 participants respectively, to assess the Nijmegen Questionnaire's effectiveness in differentiating between individuals with and without HVS (van Doorn et al 1982). This approach is considered by McDowell (2006) as an idiographic approach in item selection, which employs empirical methods to select questions that best illustrate the eventual outcome after testing a larger number of items. The professional background of these specialists (physiology, psychology, and psychiatry) was published in a different paper in the following year (van Doorn et al 1983). However, van Doorn and colleagues (1982) did not offer further details regarding the item selection process and there was no evidence to suggest the involvement of the target population in the process of content derivation, implying that their perspective is not encompassed by the measure. The Scientific Advisory Committee of the Medical Outcome Trust (2002) suggests that to meet criteria of content validity both expert and lay panels should judge the clarity, comprehensiveness, and redundancy of the items included in a measuring tool. This was only partially fulfilled by the developers of the Nijmegen Questionnaire. Considering the unavailability of this information, the level of adequacy regarding the selected items in relation to the conceptual basis of the Nijmegen Questionnaire warrants further investigation.

Furthermore, the title of the questionnaire appeared to only reflect its geographical origin (the city of Nijmegen in the Netherlands). The absence of

association between the name and content of the questionnaire potentially reduced the face validity of the Nijmegen Questionnaire, which is related to its acceptability for individuals being assessed (Bowling 1997, Laver Fawcett 2007). Thus, on the COSMIN evidence for the content validity is rated as poor (Mokkink 2010, Terwee et al 2012).

Construct validity

In the 1985 publication by van Dixhoorn and Duivenvoorden (1985), non-metric principal components analysis (NMPCA) was employed to assess the complexity of the Nijmegen Questionnaire for HVS complaints. This was the first easily identifiable step in relation to the construct validating process for the Nijmegen Questionnaire. The NMPCA was utilised to establish the dimensional structure of items included in the questionnaire and hence the structural validity (a form of construct validity) of the instrument (Tabachnick and Fidell 1996, van Dixhoorn and Duivenvoorden 1985). Three components (respiratory, central tetany, and peripheral tetany) were identified by the application of factor analysis and these followed the classic triad of HVS related complaints (Lum 1975). A key limitation of the study was an inadequate sample size to examine the structural validity of the Nijmegen Questionnaire; 75 patients were included, compared to sample size recommendations ranging between five to 10 people per item in the questionnaire (Thompson 2004).

The construct validity of the Nijmegen Questionnaire was also examined using linear analysis of discriminance (van Dixhoorn and Duivenvoorden 1985). The authors performed the analysis to establish whether the question items were able to discriminate optimally between individuals with and without HVS, hence assessment of discriminative validity (Streiner and Norman 2008). The researchers found

significant differences in the scores between the individuals with HVS and those without across all components (van Dixhoorn and Duivenvoorden 1985). In other words, participants with HVS scored distinctly higher in all three groups of complaints in the Nijmegen Questionnaire compared to those without the syndrome. Despite the appropriate application of statistical methods throughout the testing process, the quality rating on the COSMIN checklist (Mokkink 2010, Terwee et al 2012) was reduced by the inadequate sample size, omission of clear hypotheses regarding the correlations, and how missing data were managed.

Criterion validity

Some evidence to support the criterion validity of the Nijmegen Questionnaire was presented in 1983 (van Doorn et al 1983). Participants with HVS previously diagnosed by the hyperventilation provocation test (criterion/'gold standard') and those without the disease were asked to complete the Nijmegen Questionnaire and discriminant analysis was employed through the validating process. The authors summarised that the total scores of Nijmegen Questionnaire correlated strongly with the hyperventilation provocation test (van Doorn et al 1983). In addition to the inadequate sample size, the study did not provide sufficient information regarding the percentage of missing data and how this was managed, thus the evidence for the criterion validity of the questionnaire was deemed fair instead of excellent (Mokkink 2010, Terwee et al 2012). In the 1985 study, the researchers demonstrated that the Nijmegen Questionnaire possessed a greater degree of specificity (94%) than sensitivity (89%) (van Dixhoorn and Duivenvoorden 1985). This suggested that the number of false alarms or false positives (i.e. people without HVS who were identified as having HVS) was less than the number of false negatives (i.e. HVS sufferers who were incorrectly identified as healthy). The authors concluded that the Nijmegen

Questionnaire was a suitable screening tool for HVS (Bowling 2001, van Dixhoorn and Duivenvoorden 1985). It was suggested that results acquired by a screening tool (e.g. Nijmegen Questionnaire) should be subjected to a diagnostic test (e.g. Hyperventilation Provocation Test) to rule out false positives (van Doorn et al 1983).

Decisions around the cut-off point for a screening tool need to be considered in relation to specificity and sensitivity (Laver Fawcett 2007). McDowell (2006) proposed that 'if the goal is to rule out a diagnosis, a cut-off point will be chosen that enhances sensitivity, whereas if the clinical goal is to rule in a disease the cut-off point will be chosen to enhance specificity' (p 32). Although the cut-off score of 23/64 for the Nijmegen Questionnaire is documented (Garssen et al 1984, van Doorn et al 1983, Vansteenkiste et al 1991) and applied in the multidisciplinary health settings (Chaitow et al 2002), the empirical evidence that supports this is unclear in the literature. Van Doorn and colleagues (1983) was the only research team that supported their recommendation with original research. The authors suggested 22 as the cut-off score and recommended that patients who were identified with HVS to undergo the hyperventilation provocation test to rule out false positives. In the following year, Garssen and colleague (1984) suggested the currently accepted cut-off score (23/64) based on the summary of the research paper published by van Doorn and colleague (1983) without carrying out their own evaluation of patients. Although Garssen and colleague (1984) recommended how the Nijmegen Questionnaire should be administered, the credibility of this publication was diminished due to the lack of raw research data.

Cultural validity

The Nijmegen Questionnaire was developed in the Netherlands. While this questionnaire has been widely used in the field of clinical practice and health research (Chaitow et al 2002), there was no literature available for critique in terms of its cultural validity. Without subjecting this questionnaire to a recognised cultural-adaptation process, the utilisation of this tool by health professionals working in different cultural contexts could significatnly impact on clinical and research outcomes.

Reliability

The test-retest reliabilty of the Nijmegen Questionnaire was investigated by van Doorn and researchers (1983). They concluded that the questionnaire was relatively stable given the coefficient of 0.87 but, they didn't state what correlation coefficient they used prior to data testing. The authors made the decision to retain all 16 items from the Nijmegen Questionnaire based on the range of bi-serial correlations obtained (.30 to .65) indicating that all items associated with presentation of HVS. The researchers stated that the similarity between the retained symptoms of HVS was minimal based on the inter-correlations between all of the items (0.03 to 0.52) (all items captured different aspects of HVS). Evidence for the reliability of the tool was rated as fair because the authors did not report how missing data were managed and Kappa statistics were not presented (Mokkink 2010, Terwee et al 2012). Internal consistency of the tool has not been investigated to date.

Clinical utility

Clinical utility is an important factor when evaluating the quality of an assessment (Laver Fawcett 2007). An empirically validated and standardised instrument does not

automatically warrant relevance and usefulness of the tool in practice (Chaitow et al 2002). The clinical utility of an assessment tool can generally be judged in five categories: cost, time, energy and effort, portability, and acceptability (Laver Fawcett 2007).

Cost

The Nijmegen Questionnaire was published in the 1980s and it remains free for anyone to access. The ease of accessibility is evident as the content of the questionnaire is found in our literature search (van Doorn et al 1982). There is cost involved when producing copies of the test in practice but, no costly specialised training is required to administer the test.

Time

The time required for a patient to complete the Nijmegen Questionnaire is approximately five minutes (Garssen et al 1984). More time will be needed if an interpreter is required. Poor mental state and stamina resulting from an extended assessment can affect the validity and reliability of a test (Laver Fawcett 2007). In physiotherapy practice, the Nijmegen Questionnaire allows quick screening of HVS symptoms. It requires minimal preparation and results can be calculated and interpreted immediately.

Energy and effort

The energy and effort associated with the administration of an instrument is related to both the test administrator and the patient (Laver Fawcett 2007) and can influence the use of the test in health services (Chaitow et al 2002). Tests usually require less energy with repeated use (Laver Fawcett 2007). Anecdotally, the Nijmegen

Questionnaire is relatively short with 16 short questions and the administration is usually effortless in the author's (Li Ogilvie) area of practice.

Portability

The portability of an assessment tool reflects the ease of carrying or transporting an instrument (Laver Fawcett 2007). A measure that is bulky or heavy has a low portability. The Nijmegen Questionnaire can be completed as a pen and paper exercise which is highly portable.

Acceptability

The philosophy, theoretical frameworks, and interventions within a health service are to be considered when assessing the acceptability of a measure (Laver Fawcett 2007). Practitioners are encouraged to ascertain if the outcome measure is tolerated by the individuals being evaluated (Chaitow et al 2002). If a test is prone to cause distress, it might not be easily accepted by patients or their families. Patients from the lead author's clinic report that the questionnaire allows them to make sense of the symptoms of HVS and provides a baseline for progress monitoring.

Discussion

The current review identifies a small number of studies concerning the validity, reliability, and the development of the Nijmegen Questionnaire. Moreover, only two studies contained original research. Considering the limited evidence presented around three decades ago, it is remarkable how the questionnaire is still widely used in clinical and research practice. The methodological flaws that can be identified from two original research studies using the COSMIN include the lack of target population

involvement and missing items reporting, insufficient participants and statistical testing. Other measurement properties that are part of the COSMIN checklist such as internal consistency, measurment error, responsiveness, and cultural validity are not researched to date. Some of the methodolgoical flaws can be addressed by designing and carrying out studies with more participants, with the application of more robust statistical tests to generate results that can be used to better evaluate the validity and reliablity of the Nijmegen Questionnaire.

While the COSMIN checklist is a very detailed and comprehensive evaluation tool, it requires that the lowest rating to be taken as the final methodological quality score per category, i.e. the worse score counts. It means that a measurement property of the Nijmegen Questionnaire can be rated poor overall (see Table 3) despite having other questions in the same category rated higher (e.g. fair, good, or excellent). Consequently it is is important to review each COSMIN domain prior to future research so that researchers can specifically design studies that meet all the criteria for a robust study design.

While the existing evidence on validity and reliability of the measuring tool is scant, the Nijemegen Questionnaire is the only outcome measure that is suggested to be suitable for screening of hyperventilation syndrome in adults. Until further research studies are carried out to investiagte its measurement properties, reviewing the cultural validity and clinical uitily of the tool can also be meaningful.

Conclusion

This paper provides a critical summary of the validity, reliability, and clinical utility of the Nijmegen Questionnaire. The number of existing journal articles on validity and

reliability of this outcome measure is minimal. The research studies that were identified have fair to poor methodological properties. In particular, the evidence for the content validity, structural validity, and reliability is poorly represented in the studies reviewed and no research has been carried out on the cultural validity of the Nijmegen Questionnaire.

Nevertheless, the Nijmegen Questionnaire is used by health professionals as a diagnostic or screening tool for HVS (Chaitow et al 2002, Vansteenkiste et al 1991). While there is no evidence in the literature that specifically investigates the questionnaire's ability to measure change, the Nijmegen Questionnaire is often used as an outcome measure in clinical research (Agache et al 2012, Humphriss et al 2004, Thomas et al 2003). The lack of empirical evidence on the conceptual framework in relation to this instrument places doubt on the validating processes thus far. Physiotherapists who are considering or are already using this outcome measure need to be aware of the issues raised in this article when interpreting the scores and it is recommended that results gathered using the Nijmegen Questionnaire are to be interpreted in conjunction with other clinical assessments when diagnosing patients with hyperventilation. Going forward, researchers can explore and re-establish the content and conceptual basis of the Nijmegen Questionnaire by involving individuals with HVS, examine the test-retest reliability, as well as the structural and internal validity more robustly with appropriate sample sizes and statistical techniques. Until such time, there is limited evidence for the use of the only questionnaire for hyperventilation screening or diagnostic testing.

KEY POINTS

The Nijmegen Questionnaire is widely used in the screening of hyperventilation syndrome in health settings.

There is only a limited number of fair to poor quality studies evaluating the efficacy of the Nijmegen Questionnaire.

Physiotherapists and other health professionals need to be aware of the limited evidence base for this tool.

Further research that involves more robust statistical analysis is required to establish the validity, reliability, and sensitivity of the Nijmegen Questionnaire.

REFERENCES

Agache I, Ciobanu C, Paul G, Rogozea L (2012) Dysfunctional breathing phenotype in adults with asthma: Incidence and risk factors. *Clinical and Translational Allergy* 2: 18.

Beaton D, Bombardier C, Guillemin F, Ferrax M (2000) Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 25: 3186-3191.

Bowling A (1997) Measuring Health: A Review of Quality of Life Measurement Scales (2nd edn). Buckingham: Open University Press.

Bowling A (2001) Measuring Disease (2nd ed). Buckingham: Open University Press.

Chaitow L, Bradley D, Gilbert C (2002) Multidisciplinary approaches to breathing pattern disorders. Edinburgh: Churchill Livingstone.

Courtney R, Greenwood KM (2009) Preliminary investigation of a measure of dysfunctional breathing symptoms: The Self Evaluation of Breathing Questionnaire (SEBQ). *International Journal of Osteopathic Medicine* 12: 121-127.

Garssen B, Colla P, van Dixhoorn J, van Doorn P, Folgering H, Stoop A, de Swart J (1984) Het herkennen van het hyperventilatiesyndroom [Recognising the hyperventilation syndrome]. *Medisch Contact* 35: 1122-1124.

Grossman P, de Swart J (1984) Diagnosis of hyperventilation syndrome on the basis of reported complaints. *Journal of Psychosomatic Research* 28: 97-104.

Hambleton R, Jones R (1993) An NCME instructional module on comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice* 12: 39-47.

Hornsveld HK, Garssen B, Fiedeldij Dop MJ, van Spiegel PI, de Haes JC (1996) Double-blind placebo-controlled study of the hyperventilation provocation test and the validity of the hyperventilation syndrome. *The Lancet* 348: 154-158.

Høegh M, Høegh SM (2009) Trans-adapting outcome measures in rehabilitation: Cross-cultural issues. *Neuropsychological Rehabilitation* 19: 955-970.

Laver Fawcett A (2007) Principles of Assessment and Outcome Measurement for Occupational Therapists and Physiotherapists: Theory, Skills and Application. West Sussex: John Wiley and Sons.

Lum L (1975) Hyperventilation: The tip of the iceberg. *Journal of Psychosomatic Research* 19: 375-383.

McDowell I (2006) Measuring Health: A Guide to Rating Scales and Questionnaires (3rd edn). New York: Oxford University Press.

Mokkink LB (2010) The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Quality of Life Research* 19: 539-549.

Mooney S, Candy S (2008) The real cost of effective treatment: A single case study of a patient with hyperventilation syndrome. *New Zealand Journal of Physiotherapy* 36: 88.

Pallant J (2001) SPSS Survival Manual: A Step by Step Guide to Data Analysis Using SPSS. Berkshire: Open University Press.

Rapee RM, Medoro L (1994) Fear of physical sensations and trait anxiety as mediators of the response to hyperventilation in nonclinical subjects. *Journal of Abnormal Psychology* 103: 693-699.

Rowley J, Nicholls DN (2006) Development of the RoBE self-efficacy scale for people with breathing pattern disorders. *New Zealand Journal of Physiotherapy* 34: 131-141.

Scientific Advisory Committee of the Medical Outcomes Trust (2002) Assessing health status and quality-of-life instruments: Attributes and review criteria. *Quality of Life Research* 11: 193-205.

Streiner D, Norman G (2008) Health Measurement Scales: A Practical Guide to Their Development and Use (4th edn). Oxford: Oxford University Press.

Tabachnick BG, Fidell LS (1996) Using Multivariate Statistics (3rd edn). New York: Harper Collins.

Terwee CB, Mokkink LB, Knol DL, Ostelo R, Boutex LM, de Vet H (2012) Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. *Quality of Life Research* 21: 651-657.

Thomas M, McKinley R, Freeman E, Foy C, Prodger P, Price D (2003) Breathing retraining for dysfunctional breathing in asthma: A randomised controlled trial. *Thorax* 58: 110-115.

Thompson B (2004) Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications. Washington DC: American Psychologial Association.

van Dixhoorn J, Duivenvoorden H (1985) Efficacy of Nijmegen Questionnaire in recognition of the hyperventilation syndrome. *Journal of Psychosomatic Research* 29: 199-206.

van Doorn P, Colla P, Folgering H (1983) Een vragenlijst voor hyperventilatieklachten [A questionnaire for hyperventilation symptoms]. *De Psycholoog* 18: 573-577.

van Doorn P, Folgering H, Colla P (1982) Control of the end-tidal PCO_2 in the hyperventilation syndrome: Effects of biofeedback and breathing insturctions compared. *Bulletin Europeen De Physiotherpathologie Respiratoire* 18: 829-836.

Vansteenkiste J, Rochette F, Demedts M (1991) Diagnostic tests of hyperventilation syndrome. *European Respiratory Journal* 4: 393-399.

APPENDIX

Example of the Nijmegen Questionnaire

	Not at all	Rare	Sometimes	Often	Very often
Symptoms	0	1	2	3	4
Chest pain					
Feeling tense					
Blurred vision					
Dizzy spells					
Feeling confused					
Faster or deeper breathing					
Short of breath					
Tight feelings in chest					
Bloated feeling in stomach					
Tingling fingers					
Unable to breathe deeply					
Stiff fingers or arms					
Tight feelings around mouth					
Cold hands or feet					
Palpitations					
Feelings of anxiety					
					Total:

Note. The questionnaire is completed by marking how often an individual suffers from the symptoms listed. The item scores are added up to give a total score out of 64 as an indication for the presence of hyperventilation syndrome.

DISCLOSURES

Funding Source

Partial funding was granted for this paper from the Cardiothoracic Special Interest Group, Physiothearpy New Zealand.

Conflict of Interest

There was no conflict of interest and/or bias that interfered with any stage of the writing or publication process.