University of Groningen

Necessary steps in factor analysis

Schonrock-Adema, Johanna; Heijne-Penninga, Marjolein; van Hell, Elisabeth A.; Cohen-Schotanus, Janke

Published in:
Medical Teacher

DOI:
[10.1080/01421590802516756](https://doi.org/10.1080/01421590802516756)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2009

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Schonrock-Adema, J., Heijne-Penninga, M., van Hell, E. A., & Cohen-Schotanus, J. (2009). Necessary steps in factor analysis: Enhancing validation studies of educational instruments. The PHEEM applied to clerks as an example. *Medical Teacher*, 31(6), E226-E232. <https://doi.org/10.1080/01421590802516756>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

This article was downloaded by: [University of Groningen]

On: 25 October 2009

Access details: Access Details: [subscription number 906391424]

Publisher Informa Healthcare

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK

Medical Teacher

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title-content=t713438241>

Necessary steps in factor analysis: Enhancing validation studies of educational instruments. The PHEEM applied to clerks as an example

Johanna Schönrock-Adema ^a; Marjolein Heijne-Penninga ^a; Elisabeth A. van Hell ^a; Janke Cohen-Schotanus ^a

^a University of Groningen and University Medical Center Groningen, The Netherlands

First Published: June 2009

To cite this Article Schönrock-Adema, Johanna, Heijne-Penninga, Marjolein, van Hell, Elisabeth A. and Cohen-Schotanus, Janke (2009) 'Necessary steps in factor analysis: Enhancing validation studies of educational instruments. The PHEEM applied to clerks as an example', *Medical Teacher*, 31:6, e226 — e232

To link to this Article: DOI: 10.1080/01421590802516756

URL: <http://dx.doi.org/10.1080/01421590802516756>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Necessary steps in factor analysis: Enhancing validation studies of educational instruments. The PHEEM applied to clerks as an example

JOHANNA SCHÖNROCK-ADEMA, MARJOLEIN HEIJNE-PENNINGA, ELISABETH A. VAN HELL & JANKE COHEN-SCHOTANUS

University of Groningen and University Medical Center Groningen, The Netherlands

Abstract

Background: The validation of educational instruments, in particular the employment of factor analysis, can be improved in many instances.

Aims: To demonstrate the superiority of a sophisticated method of factor analysis, implying an integration of recommendations described in the factor analysis literature, over often employed limited applications of factor analysis. We demonstrate the essential steps, focusing on the Postgraduate Hospital Educational Environment Measure (PHEEM).

Method: The PHEEM was completed by 279 clerks. We performed Principal Component Analysis (PCA) with varimax rotation. A combination of three psychometric criteria was applied: scree plot, eigenvalues >1.5 and a minimum percentage of additionally explained variance of approximately 5%. Furthermore, four interpretability criteria were used. Confirmatory factor analysis was performed to verify the original scale structure.

Results: Our method yielded three interpretable and practically useful dimensions: *learning content and coaching*, *beneficial affective climate* and *external regulation*. Additionally, combining several criteria reduced the risk of overfactoring and underfactoring. Furthermore, the resulting dimensions corresponded with three learning functions essential to high-quality learning, thus strengthening our findings. Confirmatory factor analysis disproved the original scale structure.

Conclusions: Our sophisticated approach yielded several advantages over methods applied in previous validation studies. Therefore, we recommend this method in validation studies to achieve best practice.

Introduction

This study focuses on the validation of educational instruments, in particular the employment of factor analysis. In many instances, the application of factor analysis can be improved. We provide guidelines derived from literature for the proper use of factor analysis and illustrate the steps necessary to perform thorough factor analysis, focusing on educational environment instruments and, in particular, the Postgraduate Hospital Educational Environment Measure (PHEEM) (Roff et al. 2005). The validation method described in this article is also relevant for validating other educational measures, for example, instruments measuring learning styles and strategies, teaching styles, readiness for interprofessional learning, student motivation and student reflection.

The significance of the educational environment for student learning has been increasingly acknowledged in recent years (Parry et al. 2002; Miles & Leinster 2007). Over the past 10 years, many qualitative studies have been performed to discover the characteristics of the clinical educational environment. These studies have led to numerous articles describing

Practice points

- Combine several psychometric criteria in factor analysis: scree plot, eigenvalues and minimum percentage of additionally explained variance (approximately 5%).
- Tighten the eigenvalues criterion to prevent overfactoring.
- Investigate the interpretability of more than merely the solution which seems best according to the psychometric criteria.
- Relate outcomes of factor analysis to existing theories.
- Report the analysis properly.

perceptions of the clinical educational environment, and to the development of several instruments measuring the quality of the educational environment (Roff et al. 1997; Cassar 2004; Holt & Roff 2004; Roff et al. 2005; Deketelaere et al. 2006; Kanashiro et al. 2006). In most of these studies, the perceptions are clustered in themes, components or scales.

Correspondence: Johanna Schönrock-Adema, Center for Research and Innovation in Medical Education, University of Groningen and University Medical Center Groningen, A. Deusinglaan 1, 9713 AV Groningen, The Netherlands. Tel: 31 50 3638378; fax: 31 50 3637390; email: j.schonrock-adema@med.umcg.nl

Although these classifications display some overlap, consistency is lacking, perhaps because they were mainly formed on a qualitative basis. Quantitative studies applying thorough factor analysis to validate these classifications are hard to find. Often researchers adopt the original scales and examine the differences in scores on these scales between groups of respondents as quantitative validation (e.g. Pimparyon et al. 2000; Cassar 2004; Kanashiro et al. 2006). Although this is a good method to establish construct validity, factor analysis should precede this kind of validation (Tabachnick & Fidell 1996).

Factor analysis is one of the most useful methods for studying and validating the internal structure of instruments (Nunnally 1978; Pedhazur & Schmelkin 1991; Kieffer 1999; Henson & Roberts 2006). Although some researchers have applied factor analysis in validating educational environment instruments, closer inspection reveals that in most instances they did not perform or report the factor analysis thoroughly. For example, in some studies the criteria applied in factor analysis were not reported (Sobral 2004; De Oliveira Filho et al. 2005). In other studies, factor solutions were chosen on the basis of only one criterion (Nagraj et al. 2006; Boor et al. 2007; Clapham et al. 2007). The unsuitability of this method will be explained below. Furthermore, in some studies the description of the resulting factors was lacking or incomplete (Nagraj et al. 2006; Aspegren et al. 2007; Clapham et al. 2007). In addition, Nagraj et al. (2006) and Clapham et al. (2007) did not perform factor analysis *prior to* examining construct validity, but *afterwards*, which is the wrong way around (Tabachnick & Fidell 1996).

Although several criteria can be applied in factor analysis, a too limited selection of criteria can lead to widely divergent outcomes. This is illustrated by two studies in which the PHEEM was validated by applying only one criterion, with the chosen criterion being different. Clapham et al. (2007) applied the 'eigenvalues > 1' criterion, resulting in a 10-factor solution, whereas Boor et al. (2007) applied the scree test, resulting in a 1-factor solution. In both instances, the outcomes of the factor analysis are questionable: whereas the 'eigenvalue > 1' rule carries the risk of overfactoring, the 1-factor solution carries the risk of underfactoring. Overfactoring or overextraction means overestimation of the number of factors to retain. It implies that factors with little theoretical value are retained as substantial constructs and can lead to a distortion of interpretations, the development of unnecessarily complex theories and a diminishing of the replicability of the outcomes (Gorsuch 1983; Velicer & Jackson 1990; Fabrigar et al. 1999; Henson & Roberts 2006). Underfactoring is even more severe than overfactoring, as the estimated factors are likely to contain considerable error (Wood et al. 1996). Furthermore, the practical merits of a one-dimensional scale are questionable because it implies that only the overall score on the instrument is useful. An instrument containing several scales may be more practical for managing environmental change, as it may offer educators starting points for improvement by providing them with an overview of types of aspects judged as more or less positive.

Considering the divergent outcomes of applying only one criterion and the attached shortcomings, it seems better to combine several psychometric criteria in factor analysis, as

Hatcher (1994) and Kieffer (1999) recommended. Hatcher (1994) recommends applying three psychometric criteria together: the scree test, the eigenvalues criterion, and the proportion of variance that a factor explains. The factor solution that seems best according to all these criteria should be subjected to a further investigation into the meaningfulness of the factors. Investigating interpretability is essential, as a model that fails to produce a rotated solution that is interpretable and theoretically sensible is of little value (Rummel 1977; Fabrigar et al. 1999). However, merely applying psychometric criteria does not necessarily lead to the best interpretable solution. To ensure selection of the solution that makes the most sense and displays the most scientific sensibility, researchers suggest investigating the interpretability of more than merely that solution which seems best according to the psychometric criteria (Tabachnick & Fidell 1996; Lee & Hooley 2005).

The aim of this study was to demonstrate the superiority of a sophisticated method, implying an integration of the recommendations described in the factor analysis literature, over more limited methods. For this demonstration we validated the PHEEM, employing a combination of several psychometric criteria and investigating the interpretability of the resulting factor solutions. Proving that our approach is better than previously applied methods implies that it yields (a) interpretable and practically useful dimensions with (b) less risk of under and overfactoring and (c) theoretically sensible dimensions that are more in accordance with educational theories than the outcomes in the studies mentioned before. In addition, we investigated how this classification relates to the original scaling of the PHEEM.

Methods

Respondents and procedure

Data were collected at the University Medical Center of Groningen. Participation was voluntary. The PHEEM was anonymously completed by 279 clerks from 8 hospitals (68%) before sitting a progress test.

Instrument

The PHEEM consists of 40 items on a Likert-type scale (1 = totally disagree, 5 = totally agree). We used the Dutch version of the PHEEM developed and validated by Boor et al. (2007). Their translation of the PHEEM into Dutch was completed with the original authors' permission. A professional translator rendered this Dutch version back into English. The original authors considered this version equivalent to the original questionnaire. Although some items were adapted to the Dutch situation, for example, item 17, 'My hours conform to the New Deal', the substance of the items did not differ from that of the original items. For reasons of recognizability, we decided to use the original formulations of the items in our results. Three scales were originally identified: perceptions of (a) role autonomy (15 items), (b) teaching (14 items) and (c) social support (11 items) (Roff et al. 2005).

Analysis

Before performing the factor analysis, we checked whether the assumptions regarding normality of the distribution were satisfied. Skewness and kurtosis are used as indicators of the normality among single variables (Hatcher 1994; Tabachnick & Fidell 1996). In case of substantial skewness and kurtosis, which may degrade the solution, variable transformation is considered (for details see Tabachnick and Fidell 1996, pp. 82–85). Two items (items 7 and 13) were transformed because they did not satisfy the assumptions of normality. Because of the sample size, which was good (Tabachnick & Fidell 1996, p. 640) and the distribution of the scores, we assumed an interval level, which is a tenable assumption (Jaccard & Wan 1996), and applied parametric statistics.

To investigate the internal structure of the PHEEM, we applied Principal Components Analysis with varimax rotation. Several criteria were applied to determine how many factors should be retained (Table 1):

- (1) The point of inflexion displayed by the scree plot;
- (2) The eigenvalues criterion. Since several studies show that the ‘eigenvalues > 1’ rule leads to an overestimation of the number of factors to retain (Henson & Roberts 2006), in this study this rule was tightened to ‘eigenvalues > 1.5’;
- (3) The ‘proportion of variance accounted for’ criterion. A component was retained if it minimally explained an approximate additional 5% of the variance.

The interpretability was investigated using Hatcher’s interpretability criteria (Hatcher 1994). To find the best solution in terms of interpretability and theoretical sensibility, we decided to investigate the interpretability of the best solution according to the current three psychometric criteria and that of solutions with up to two factors more and two factors less (Lee & Hooley 2005). The interpretability criteria read:

- (4a) A given component contains at least three variables with significant loadings, a loading of 0.40 being suggested as the cut-off point;
- (4b) Variables loading on the same component share the same conceptual meaning;
- (4c) Variables loading on different components appear to measure different constructs;

(4d) The rotated factor pattern demonstrates ‘simple structure’, which means that:

- i most variables load relatively high on only one component and low on the other components;
- ii most components have relatively high factor loadings for some variables and low loadings for the remaining ones.

The interpretation process for each rotated solution started with the elimination of double-loading items (criterion 4d), which are items that load at least 0.40 on more than one factor (Hatcher 1994). After ascertaining whether the solution satisfied all the interpretability criteria, the analyses were rerun without the double-loading items.

To determine whether criteria 4b and 4c were satisfied, eight medical education experts independently interpreted the factors for all relevant solutions. Subsequently, the interpretations were discussed and the expert team reached complete agreement on (a) the interpretations of the factors and (b) the best factor solution.

A *confirmatory factor analysis*, namely, the Oblique Multiple Group Method (OMG), was performed to verify whether the original PHEEM structure should be maintained (Stuive et al. 2008). The first step in the OMG involves constructing scales by simply combining the items assigned to the same scale. In the next step the correlation of each item with each scale is calculated. Items should correlate strongest with the scale to which they are assigned. If items correlate stronger with other scales, this indicates they were wrongly assigned. The strongest correlation indicates the scale the item should be assigned to instead.

Results

Principal Component Analysis

The data displayed several missing data randomly scattered over the dataset. Because listwise exclusion would result in a sample size too small for reliable factor analysis and pairwise deletion carries the risk of less replicable outcomes, we performed the analysis replacing missing values with the variable mean (Field 2006). The scree plot showed a sharp point of inflection (criterion 1) after the first factor (Figure 1). Only five factors had initial eigenvalues > 1.5 (criterion 2), with values ranging from 1.55 to 10.98. Of these, only the first three factors accounted for more than or approximately 5% of the variance (criterion 3). Considering the eigenvalue and the ‘proportion of variance accounted for’ criterion, the 3-factor solution was taken as the starting point for our analysis. Our previous decision to investigate the interpretabilities of solutions with up to two factors more and two factors less to find the best interpretable solution, was revised. For obvious reasons, we decided to drop the 1-factor solution from further investigation, which left the 2-, 3-, 4- and 5-factor solutions.

After eliminating items loading on more than one factor (criterion 4d), all the factors of each solution still contained at least three items with significant loadings (criterion 4a). However, the 4- and 5-factor solutions contained factors that were not unambiguously interpretable (criterion 4b).

Table 1. Psychometric and interpretability criteria applied in factor analysis.

Criteria		
Objective or psychometric criteria	1	Point of inflexion on the scree plot
	2	Eigenvalues > 1.5
	3	Proportion of variance accounted for is minimally approximately 5%
Interpretability criteria	4a	At least 3 variables with a loading >0.40 per factor
	4b	Variables of the same component measure the same construct
	4c	Variables loading on different components measure different constructs
	4d	The rotated factor pattern demonstrates simple structure

Figure 1. Scree plot of the eigenvalues of the factors.

In addition, they failed to meet the criterion that different factors should measure different constructs (criterion 4c). Of the remaining two solutions, the 3-factor solution was easier to interpret than the 2-factor solution (criterion 4b) – the first factor of the 2-factor solution seemed to measure two different constructs which were separated in the 3-factor solution. The 3-factor solution explained 37.7% of the variance. As this solution initially had three items loading on two factors, namely, (21) 'There is access to an educational programme relevant to my needs', (31) 'My clinical teachers are accessible' and (40) 'My clinical teachers promote an atmosphere of mutual respect', the analyses were repeated without these items. This resulted in similar outcomes, with one item again loading on two factors – (10) 'My clinical teachers have good communication skills'. After removal of this item the analyses were rerun. The resulting 3-factor solution demonstrated a perfectly 'simple structure', with interpretations that did not differ from the initial labels assigned to the factors.

The final, 3-factor solution explained 37.1% of the variance, with the third, unrotated factor explaining 5.0%. The initial eigenvalues of the components were 9.42, 2.14 and 1.78. After rotation, the factors explained 21.95%, 7.63% and 7.48% of the variance. The first scale (Table 2) concerned aspects related to the content of the clerkship and the quality of coaching by teachers. The experts interpreted this factor as *learning content and coaching*. The second factor contained aspects concerning the social climate, also interpreted as culture or atmosphere. This factor was summarized as *beneficial affective climate*. The third factor mainly focused on the organization of the clerkship – providing necessary conditions for educational activities – which could also be labelled as educational or *external regulation*. The reliability of the 36 item PHEEM was high, with a Cronbach's α of 0.91. The scales had α 's of 0.92, 0.61 and 0.66, respectively. The third factor consisted of only five items and correction for test length

showed that on an 11-item scale, which is the length of the shortest scale in the original scale structure (Roff et al. 2005), the α would increase to 0.81.

Quality criteria. In the Section 'Introduction' we stated, proving that our approach is better than more limited methods implies that it yields (a) interpretable and practically useful dimensions with (b) less risk of under and overfactoring and (c) theoretically sensible dimensions that are more in accordance with educational theories. First, the three factors that we found were interpretable and appear to be practically useful. Second, our findings fall between the outcomes of the aforementioned studies applying only one criterion (Boor et al. 2007; Clapham et al. 2007), thus carrying less risk of overfactoring and underfactoring. This benefit can be ascribed to the more sophisticated approach of combining several criteria, considering the fact that our results were comparable to those of the PHEEM studies mentioned previously (Boor et al. 2007; Clapham et al. 2007): if we had applied the same criteria as used in those studies, we would have found 10 factors with eigenvalues > 1 versus 1 factor with the point of inflexion criterion. Third, our dimensions appeared to be theoretically sensible, as they corresponded closely with theories on learning functions essential to high-quality learning (Shuell 1988; Vermunt 1996; Vermunt & Verloop 1999). In these theories that focus on theoretical learning, three main types of learning functions are identified that are important for enhancing the learning process: cognitive, affective and metacognitive functions. Our first dimension, *learning content and coaching*, corresponds with the cognitive learning functions, which pertain to the acquisition of learning content and lead directly to learning outcomes (Vermunt & Verloop 1999). Due to the different setting (workplace learning rather than theoretical learning), the emphasis in our dimension is more on adequate learning opportunities and coaching of learning content than on

Table 2. Factor loadings and communalities of the final 3-factor solution^a.

No.	Item	Component			h ²
		1	2	3	
27	I have enough clinical learning opportunities for my needs	0.793	-0.020	-0.065	0.63
34	The training in this post makes me feel ready to be an SpR/consultant	0.752	-0.017	0.126	0.58
35	My clinical teachers have good mentoring skills	0.721	0.142	0.198	0.58
30	I have opportunities to acquire the appropriate practical procedures for my grade	0.685	-0.057	0.009	0.47
28	My clinical teachers have good teaching skills	0.678	0.170	0.156	0.51
36	I get a lot of enjoyment out of my present job	0.661	0.169	0.111	0.48
39	The clinical teachers provide me with good feedback on my strengths and weaknesses	0.660	0.092	0.158	0.47
29	I feel part of a team working here	0.658	0.108	0.055	0.45
6	I have good clinical supervision at all times	0.646	0.206	0.077	0.47
22	I get regular feedback from my seniors	0.636	0.190	0.213	0.49
15	My clinical teachers are enthusiastic	0.621	0.305	0.131	0.50
37	My clinical teachers encourage me to be an independent learner	0.607	0.012	0.159	0.40
33	Senior staff utilise learning opportunities effectively	0.601	0.150	0.253	0.45
2	My clinical teachers set clear expectations	0.581	0.123	0.299	0.44
5	I have the appropriate level of responsibility in this post	0.579	-0.037	0.305	0.43
23	My clinical teachers are well organised	0.542	0.264	0.318	0.46
32	My workload in this job is fine	0.467	0.237	0.140	0.29
19	I have suitable access to careers advice	0.419	0.266	-0.005	0.25
14	There are clear clinical protocols in this post	0.351	0.034	0.242	0.18
18	I have the opportunity to provide continuity of care	0.345	0.173	0.077	0.16
38	There are good counselling opportunities for junior doctors who fail to complete their training satisfactorily	0.313	0.052	0.307	0.20
7	<i>There is racism in this post</i>	0.075	0.661	-0.061	0.45
13	<i>There is sex discrimination in this post</i>	0.045	0.640	-0.041	0.41
11	<i>I am beeped inappropriately</i>	0.081	0.591	-0.088	0.36
8	<i>I have to perform inappropriate tasks</i>	0.155	0.550	-0.054	0.33
25	There is a no-blame culture in this post	0.253	0.416	0.115	0.24
17	My hours conform to the New Deal	-0.019	0.410	0.183	0.20
20	This hospital has good quality accommodation for junior doctors, especially when on call	0.056	0.386	0.255	0.22
26	There are adequate catering facilities when I am on call	0.031	0.306	0.161	0.12
24	I feel physically safe within the hospital environment	0.092	0.201	0.115	0.06
16	I have good collaboration with other doctors in my grade	0.163	0.180	0.128	0.08
3	I have protected time at this post	0.095	-0.007	0.678	0.47
9	There is an informative junior doctors handbook	0.097	0.098	0.655	0.45
1	I have a contract of employment that provides information about hours of work	0.131	0.037	0.564	0.34
4	I had an informative introduction programme	0.347	0.050	0.546	0.42
12	I am able to participate actively in educational events	0.164	0.188	0.513	0.33
	Initial eigenvalues	9.4	2.1	1.8	
	Explained variance after rotation	21.95%	7.63%	7.48%	

Notes: h² communality estimates (estimates of the proportion of variance in a given variable explained by all components jointly).

^aThe scores on the *italic* items were reversed.

teaching in the sense of presenting and explaining learning content. The *beneficial affective climate* dimension corresponds with the affective learning functions. These functions refer to the ability to cope with feelings arising during learning, which in turn lead to mood states that may foster or inhibit the learning process. Our dimension reflects aspects that typify the atmosphere of the educational environment and that may affect students' moods. Our *external regulation* dimension corresponds with the metacognitive learning functions that aim to regulate learning and mood states, and lead indirectly to learning outcomes by directing the course and outcomes of the learning process. The emphasis in this dimension differs most from the corresponding learning function: it mainly encompasses organizational aspects, and although these are external regulatory aspects, most aspects representative of this dimension cannot be regarded as metacognitive aspects.

Confirmatory factor analysis

The OMG showed that 18 of the total 40 items loaded higher on another scale than the scale proposed by the authors, e230

namely, 11 items from 'perceptions of autonomy' and 7 from 'perceptions of social support'. In addition, the proposed PHEEM structure explained 33.2% of the variance, 3.9% less than the variance explained by the 3-factor solution that we found using Principal Component Analysis. From these results, we conclude that the OMG did not support the proposed PHEEM structure.

Discussion

The aim of this study was to demonstrate how to validate an educational instrument taking into account the findings and recommendations described in the factor analysis literature. First, the application of several criteria in combination, as recommended by experts (Hatcher 1994; Tabachnick & Fidell 1996), led to three interpretable and practically useful dimensions of the PHEEM, *learning content and coaching*, *beneficial affective climate* and *external regulation*. Second, using our sophisticated method reduced the risk of overfactoring and underfactoring compared to the more limited methods applied in aforementioned studies (Boor et al. 2007;

Clapham et al. 2007). Last, but not least, our dimensions appeared to be theoretically sensible as they corresponded with educational theories. As the authors of the before-mentioned studies did not connect their outcomes to educational theories and we were not able either to connect their outcomes to educational theories, our outcomes seem to be theoretically more sensible, which is again in favour of our method. Replication studies focusing on other instruments are needed to prove the superiority of combining several criteria compared to less sophisticated approaches.

One limitation of this study is that the amount of explained variance is only moderate. However, we explained more variance than we would have, had we remained with the 1-factor solution of Boor (2007) or with the original scaling as proposed by Roff et al. (2005). Retaining more factors was not an option, as the interpretability criteria were violated. The strengths of our method include a reduced risk of over and underfactoring than in previous studies and the identification of factors that are interpretable, practically useful and correspond with learning theories. Therefore, we recommend applying the current combination of psychometric and interpretability criteria in future validation studies. Further research should investigate whether our current findings can be replicated.

A second possible limitation of this study is that it was based on clerks' perceptions, whereas the instrument was originally created for postgraduates. However, we consider it unlikely that the factorial structure would differ much for both samples as the workplace is essentially the same: it concerns adult learning in a clinical learning environment. In fact, the outcomes based on clerk samples seem very similar to those based on resident samples in view of previous studies and the current study (Boor et al. 2007; Clapham et al. 2007). In addition, the fact that our dimensions correspond with theories on learning functions important for theoretical learning makes it likely that the learning functions are important for learning in general. This strengthens our assertion that there is no reason to assume that these dimensions do not apply in postgraduate learning.

Although this study proves that choices made during data analysis have direct outcomes, it must also be noted that the quality of the outcomes depends partly on the quality of the instrument. From our results, we can infer that Roff et al. (2005) did an outstanding job in performing their extensive qualitative study and in reducing their item pool to the core items. The fact that the learning functions are reflected in our PHEEM data validates their work. On the other hand, this finding also supports theories on learning functions and their importance for high-quality learning.

In conclusion, this study shows several advantages of our approach over the methods applied in the previous validation studies. Therefore, we recommend that researchers aiming to validate educational instruments

- i. combine several psychometric criteria – namely, (a) scree plot, (b) eigenvalues > 1.5 and (c) components explain minimally an approximate additional 5% of the variance;
- ii. explore the interpretability of the resulting solution and of solutions with one and two factors more and less and
- iii. relate the outcomes to existing theories.

In addition, it is important to report the procedure and outcomes of the factor analysis properly. We hope that our research will stimulate the use of the current method in validation studies and that these validation studies will lead, eventually, to more consistency in the resulting dimensions found and, as a result, to best practice in educational research.

Acknowledgements

The authors would like to express their gratitude to the students who participated in this study and to Tineke Bouwkamp for her constructive comments on the manuscript.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the article.

Notes on contributors

JOHANNA SCHÖNROCK-ADEMA is a Researcher, center for research and Innovation in Medical Education, University of Groningen and University Medical Center Groningen, The Netherlands.

MARJOLEIN HEIJNE-PENNINGA is a psychologist, Institute for Medical Education, University of Groningen and University Medical Center Groningen, The Netherlands.

ELISABETH A. VAN HELL is an Educationalist, Center for Research and Innovation of Medical Education, University of Groningen and University Medical Center Groningen, The Netherlands.

JANKE COHEN-SCHOTANUS is an Associate professor and head of the Center for Research and Innovation in Medical Education, University of Groningen and University Medical Center Groningen, The Netherlands.

References

- Aspegren K, Bastholt L, Bsted KM, Bonnesen T, Ejlersen E, Fog I, Hertel T, Kodall T, Lund J, Madsen JS, et al. 2007. Validation of the PHEEM instrument in a Danish hospital setting. *Med Teach* 29:504–506.
- Boor K, Scheele F, Van der Vleuten CPM, Scherpbier AJJA, Teunissen PW, Sijtsma K. 2007. Psychometric properties of an instrument to measure the clinical learning environment. *Med Educ* 41:92–99.
- Cassar K. 2004. Development of an instrument to measure the surgical operating theatre learning environment as perceived by basic surgical trainees. *Med Teach* 26:260–264.
- Clapham M, Wall D, Batchelor A. 2007. Educational environment in intensive care medicine – use of Postgraduate Hospital Educational Environment Measure (PHEEM). *Med Teach* 29:e184–e191.
- Deketelaere A, Kelchtermans G, Struyf E, De Leyn P. 2006. Disentangling clinical learning experiences: An exploratory study on the dynamic tensions in internship. *Med Educ* 40:908–915.
- De Oliveira Filho GR, Vieira JE, Schonhorst L. 2005. Psychometric properties of the Dundee Ready Educational Environment Measure (DREEM) applied to medical residents. *Med Teach* 27:343–347.
- Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ. 1999. Evaluating the use of exploratory factor analysis in psychological research. *Psychol Methods* 4:272–299.
- Field A. 2006. *Discovering statistics using SPSS*. 2nd ed. London: Sage Publications.
- Gorsuch RL. 1983. *Factor analysis*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hatcher L. 1994. *A step-by-step approach to using the SAS system for factor analysis and structural equation modeling*. Cary, NC: SAS Institute Inc.

- Henson RK, Roberts JK. 2006. Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educ Psychol Meas* 66:393–416.
- Holt MC, Roff S. 2004. Development and validation of the Anaesthetic Theatre Educational Environment Measure (ATEEM). *Med Teach* 26:553–558.
- Jaccard J, Wan CK. 1996. LISREL Approaches to interaction effects in multiple regression. Thousand Oaks: Sage Publications.
- Kanashiro J, McAleer S, Roff S. 2006. Assessing the educational environment in the operating room – a measure of resident perception at one Canadian institution. *Surgery* 139:150–158.
- Kieffer KM. 1999. An introductory primer on the appropriate use of exploratory and confirmatory factor analysis. *Res Schools* 6:75–92.
- Lee N, Hooley G. 2005. The evolution of ‘classical mythology’ within marketing measure development. *Eur J Mark* 39:365–385.
- Miles S, Leinster SJ. 2007. Medical students’ perceptions of their educational environment: Expected versus actual perceptions. *Med Educ* 41:265–272.
- Nagraj S, Wall D, Jones E. 2006. Can STEEM be used to measure the educational environment within the operating theatre for undergraduate medical students? *Med Teach* 28:642–647.
- Nunnally JC. 1978. Psychometric theory. 2nd ed. New York: McGraw-Hill.
- Parry J, Mathers J, Al-Fares A, Mohammad M, Nandakumar M, Tsivos D. 2002. Hostile teaching hospitals and friendly district general hospitals: Final year students’ views on clinical attachment locations. *Med Edu* 36:1131–1141.
- Pedhazur EJ, Schmelkin LP. 1991. Measurement, design, and analysis: An integrated approach. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Pimparyon P, Roff S, McAleer S, Poonchai B, Pemba S. 2000. Educational environment, student approaches to learning and academic achievement in a Thai nursing school. *Med Teach* 22:359–364.
- Roff S, McAleer S, Harden RM, Al-Qahtani M, Ahmed AU, Deza H, Groenen G, Pimparyon P. 1997. Development and validation of the Dundee Ready Education Environment Measure (DREEM). *Med Teach* 19:295–299.
- Roff S, McAleer S, Skinner A. 2005. Development and validation of an instrument to measure the postgraduate clinical learning and teaching educational environment for hospital-based junior doctors in the UK. *Med Teach* 27:326–331.
- Rummel RJ. 1977. Applied factor analysis. Evanston: Northwestern University Press.
- Shuell TJ. 1988. The role of the student in learning from instruction. *Contemp Educ Psychol* 13:276–295.
- Sobral DT. 2004. Medical students’ self-appraisal of first-year learning outcomes: Use of the course valuing inventory. *Med Teach* 26:234–238.
- Stuive I, Kiers HAL, Timmerman ME, Ten Berge JMF. 2008. The empirical verification of an assignment of items to subtests: The Oblique Multiple Group Method versus the Confirmatory Common Factor Analysis. *Educ Psychol Meas* [E-publication ahead of print].
- Tabachnick BG, Fidell LS. 1996. Using Multivariate Statistics. 3rd ed. New York: Harper & Collins College Publishers.
- Velicer WF, Jackson DN. 1990. Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multiv Behav Res* 25:1–28.
- Vermunt JD. 1996. Metacognitive, cognitive and affective aspects of learning styles and strategies: A phenomenographic analysis. *Higher Educ* 31:25–50.
- Vermunt JD, Verloop N. 1999. Congruence and friction between learning and teaching. *Learn Instruc* 9:257–280.
- Wood JM, Tatarzyn DJ, Gorsuch RL. 1996. Effects of under- and over-extraction on principal axis factor analysis with varimax rotation. *Psychol Methods* 1:354–365.