

The Questionnaire D-RECT German: Adaptation and testtheoretical properties of an instrument for Evaluation of the learning climate in medical specialist training

Abstract

Aim: Boor et al [1] developed and validated the questionnaire D-RECT (Dutch Residency Educational Climate Test) to measure the clinical learning environment within the medical specialist training. In this study, a German version of this questionnaire (D-RECT German) is analyzed regarding testtheoretical properties.

Problem: Are the results of Boor et al replicable as a proof for validity of the questionnaire D-RECT?

Material & Methods: The study was performed as online survey using the questionnaire D-RECT German (50 items in 11 subscales). To determine item characteristics and internal consistency (Cronbach's α), item- and reliability analyses were performed. Furthermore, a confirmatory factor analysis was performed using a model for maximum-likelihood estimation to evaluate validity.

Results: This replication study on the psychometric properties of the D-RECT with 255 residents at 17 German hospitals revealed heterogeneous discriminatory power for all items and an internal consistency of Cronbach's α between 0.57 and 0.85. Within the confirmatory factor analysis, 6 items showed standardized regression coefficients <0.5 , two of them in the subscale "Attendings role". Furthermore, strong interdependencies (>0.7) were found between the subscales "Supervision", "Coaching" and "Attendings role".

Conclusion: The present replication study with the D-RECT German showed structural differences with respect to factorial validity underpinning the need of further validation studies.

Keywords: D-RECT, learning climate, further medical education, residents, validation, replication study

Peter Iblher^{1,2}

M. Zupanic^{2,3}

T. Ostermann^{4,5}

1 University of Lübeck, Clinic for Anaesthesiology and Intensive Care Medicine, Lübeck, Germany

2 Witten/Herdecke University, Institute for Teaching and Educational Research in Health Sciences, Witten, Germany

3 Witten/Herdecke University, School of Medicine, Student Dean's Office, Witten, Germany

4 Witten/Herdecke University, Department of Psychology, Chair for Statistics and doctrine of methodology, Witten, Germany

5 Witten/Herdecke University, Institute for Integrative Medicine, Witten, Germany

1. Background

Studying medicine has been subject to a considerable change within the last years. Clear and decisive efforts have been made to improve education of young physicians by means of medical didactics, leading to a changing perception of the priority of excellent teaching. This manifests in a variety of funding activities [2].

The license to practice medicine (approbation) as completion of this phase of education forms the initial basis for further qualification of the students with the task to enhance their skills and knowledge; and consequently to receive a specialist certification. In this context the demands regarding clinical competence and learning performance, which as a matter of course are consented as an imperative for medical education of students [3], [4], should also be applied for further medical education. Thus a sound and comprehensive specialist training has

been explicitly expressed by the German Medical Association (http://www.bundesaerztekammer.de/fileadmin/user_upload/downloads/20130628-MWBO_V6.pdf, accessed 21.10.2015) and should be scrutinized, to enable the educators to interpret their status quo and to consequently put modifications into practice and reassess them again. A central basis for this approach is given by the various forms of learning atmosphere [5]. Of note it would be desirable to assess the impact of learning climate by means of a flexible, established and reliable instrument, which maps the relevant dimensions of the learning atmosphere, unfolds strengths and weaknesses of further education concepts and which also could be applied within educational research.

The D-RECT questionnaire to evaluate residents' learning climate

Based on qualitative studies on the establishment of an optimal learning climate in further medical education the working group of Boor et al. found the following three interacting domains:

1. Working environment,
2. Further education,
3. Needs of residents.

Within this theoretical construct an instrument to measure the clinical learning environment was developed which contains eleven categories (Dutch Residency Educational Climate Test/ D-RECT) [1] (see also attachment 1). The authors concluded further investigations with respect to the validity of the questionnaire for the use in international settings would be necessary. Up to now the factorial validity of the final questionnaire still remains to be analyzed. The original questionnaire D-RECT was translated by Boor et al into English for publication reasons.

Thus, our replication study investigated the application of the D-RECT in the German-speaking area, framing the following questions:

1. Examination of the test-theoretical properties of a German Version of the D-RECT (D-RECT-German) to investigate learning climate by means of item- and reliability analyses to derive its internal consistency (Cronbach's α).
2. Investigation of factorial validity of the D-RECT-German by means of confirmatory factor analysis.

2. Material and Methods

Our study received a positive vote from the ethics-committee of the University of Witten/Herdecke. All participants were informed prior to the study and were given the opportunity to withdraw from the study. Agreement of participation was given on the basis of conclusive conduct.

2.1. The D-RECT questionnaire

The D-RECT consists of 50 items in 11 subscales (see attachment 1) using a five-point Likert scale (1=does not apply – 5= totally applies). The English version of the D-RECT was translated by a bilingual native speaker from English to German and back again. In addition to the original items, information on age, gender, specialist area, hospital and training year were inquired. The study was performed as online survey using Lime Survey. All collected data was anonymized before statistical analysis.

2.2. Test theoretical validation of the D-RECT-German

To determine item characteristics, item means (M), standard deviation (SD) and discrimination (r_{it}) were cal-

culated. Discrimination is calculated as the correlation between the question score and the overall assessment (item-total-correlation) for each of the 11 subscales [6]. Item-total correlations between 0.4 and 0.7 are considered as good, between 0.2-0.4 as acceptable, between 0.1-0.2 as marginal and between 0 and 0.1 as unsatisfactory [7]. To determine scale characteristics and scale intercorrelations, scale means (M), standard deviation (SD), coefficients of homogeneity (Cronbach's α) and corrected inter-scale-correlations (r according to Pearson) were calculated. For group comparisons a Cronbach's α of 0.7 can be regarded as satisfactory; a value of Cronbach's α greater than 0.8 as good [8]. In addition, Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test for sphericity were calculated to determine whether a confirmatory factor analysis was warranted.

Finally structural equation modeling was applied to determine, the amount of interdependency between items and constructs using the existing factorial solution as a model for maximum-likelihood estimation. The Chi-Square value served as a parameter for model validity. A significant Chi²-Test indicates a poor model fit. In addition common incremental measures of scale fit in structural, equation modeling like the Comparative Fit Index (CFI), Tucker Lewis Index (TLI) und Root Mean Square Error of Approximation (RMSEA) were calculated. For CFI and TLI values greater than 0.90 point to a good model-fit, while, a RMSEA>0.08 indicates a high amount of unexplained variance. Associations between items and given dimensions were expressed using standardized regression coefficients r_s , which in case of $r_s < 0.5$ were judged as unsatisfactory. Correlations between the dimensions were determined by correlation coefficients from the estimated covariance matrix. Correlations>0.7 pointed towards an interdependency of the factors. Using the R-procedure *modindices*, possible source of misspecification of the model were identified. Due to the low sample size no further multi-level analyses were carried out. All calculations were run with SPSS 22, AMOS 20 and R.

3. Results

Our sample included 255 residents (female: n=129/50.6%; male: n=126/49.4%) of 17 German hospitals (see Table 1) and from four medical areas (see Table 2) with a mean age of 32±6 years. Differentiation with respect to years of medical training is given in table 3.

Table 1: Differentiation of respondents with respect to federal states and locations (n=2: missing).

Federal state	Number locations	Number respondents
Schleswig-Holstein	6	104
Mecklenburg-Western Pomerania	3	39
North- Rhine- Westphalia	3	33
Berlin	2	40
Bavaria	1	15
Hamburg	2	22
total	17	253

Table 2: Differentiation of respondents with respect to specialties.

Specialty	Frequency	Percent
Anaesthesiology	198	77,6
Internal Medicine	36	14,1
Pediatric surgery	13	5,1
Family medicine	8	3,1

Table 3: Differentiation of respondents with respect to years of medical training.

Years of medical training	Frequency	Percent
1.-2. years	93	36,5
3.-4. years	80	29,8
5.-6. years	52	20,4
7.-9. years	23	9,0
>9. years	10	4,0

Results of item analysis are provided in attachment 1. High agreements with mean values higher than 4.0 were given in Item 3 ("It is clear which attending supervises me."; M: 4.3±0.9), Item 35 ("When I need a attending, I can always contact one."; M: 4.4±0.8) and Item 36 ("When I need to consult a attending, they are readily available."; M: 4.2±0.8). Lowest agreement was given for Item 10 ("My attendings occasionally observe me taking a history."; M: 1.4±0.8), Item 13 ("Observation forms (i. e. Mini-CEX) are used to structure feedback."; M: 1.1±0.6), Item 14 ("Observation forms (i. e. Mini-CEX) are used periodically to monitor my progress."; M: 1.1±0.6), Item 44 ("In this rotation evaluations are useful discussions about my performance."; M: 1.6±1.4), Item 45 ("My plans for the future are part of the discussion."; M: 1.7±1.5) and Item 46 ("During evaluations, input from several attendings is considered."; M: 1.3±1.3) with mean values lower than 2.0.

All items showed a satisfying discriminatory power with none of the items being below the critical value of 0.2 (see attachment 1).

Results of the scale analysis found high agreement of the residents with the subscale "Patient sign out" (mean±SD: 4.1±0.9). Lowest congruence was given for the scales "Feedback" (mean±SD: 1.5±0.5) and "Role of the specialty tutor" (mean±SD: 1.7±1.1). With respect to internal consistency, all subscales only showed a critical to moderate values of Cronbach's α between 0.57 und 0.85.

Examination of factorial structure of the D-RECT using confirmatory factor analysis resulted in an unsatisfactory model-fit with a highly significant chi-square value of 2383,576 ($p < 0.001$). Moreover incremental-fit parameters CFI (0.768) and TLI (0.746) by no means reached the area of a good approximative model-fit. Only the absolute model-fit RMSEA of 0.068 revealed a sufficient matching of the postulated factorial structure with the empirical data.

Correlations of the items and the preset dimensions of the structural equation model resulted in values between 0.225 and 0.957 implying a considerable heterogeneity. Six of the items showed standardised regression coefficients lower than 0.5 with the preset dimensions, of which two were found for the subscale "Attendings' role". Sufficient factor loadings were given for the dimensions "Supervision" (r_s between 0.570 and 0.720), "Teamwork" (r_s between 0.598 and 0.716), "Professional relations between attendings" (r_s between 0.596 and 0.700), "Formal education" (r_s between 0.557 and 0.842), "Role of the specialty tutor" (r_s between 0.531 and 0.817) and "Patient sign out" (r_s between 0.596 and 0.780). One loading below 0.5 was found in the dimensions "Coaching and assessment" (CA7: $r_s = 0.416$), "Feedback" (FB1: $r_s = 0.225$), "Peer collaboration" (PC3: $r_s = 0.441$), and "Work is adapted to residents' competence" (WA3: $r_s = 0.294$). Two loadings below 0.5 were found for the dimension "Attendings' role" (AR3: $r_s = 0.470$ and AR8: $r_s = 0.359$) (see Table 4). This result is also confirmed by the analysis of misspecifications of the model. Here once again the item FB1 parallelly loads on seven different dimensions and has to be considered as critical for the German Version of the D-RECT.

Correlation analysis between the dimensions with one exception ("Teamwork" vs. "Role of the specialty tutor" $r = -0.025$) found positive correlations between the scales. In particular the subscale "Coaching and assessment" revealed the highest correlations with "Attendings' role" ($r = 0.788$) and "Feedback" ($r = 0.752$). But also the scales "Attendings' role" and "Work is adapted to residents' competence" showed a critically high correlation of $r = 0.602$. All other scales correlated in an acceptable discriminatory range below 0.6. Subscale 11 ("Attendings' role") exhibited the highest correlations with the other subscales (see Table 5).

4. Discussion

This replication study on the psychometric properties of the D-RECT revealed acceptable to good discriminatory

Table 4: Factor loadings (standardized regression coefficients r_s between items and dimensions). Abbreviations see attachment 1.

	SV	CB	FB	TW	ZWB	PBÄ	FWB	EBF	FA	RFA	PÜ
SV1	0.686										
SV2	0.720										
SV3	0.581										
CB1		0.517									
CB2		0.646									
CB3		0.691									
CB4		0.772									
CB5		0.694									
CB6		0.665									
CB7		0.416									
CB8		0.581									
FB1			0.225								
FB2			0.905								
FB3			0.957								
TW1				0.598							
TW2				0.713							
TW3				0.716							
TW4				0.608							
ZW1					0.640						
ZW2					0.698						
ZW3					0.441						
PB1						0.586					
PB2						0.700					
PB3						0.596					
FW1							0.602				
FW2							0.646				
FW3							0.294				
FW4							0.641				
EB1								0.783			
EB2								0.797			
EB3								0.470			
EB4								0.658			
EB5								0.690			
EB6								0.615			
EB7								0.609			
EB8								0.359			
FA1									0.557		
FA2									0.749		
FA3									0.842		
FA4									0.815		
RF1										0.554	
RF2										0.531	
RF3										0.603	
RF4										0.800	
RF5										0.782	
RF6										0.817	
PÜ1											0.596
PÜ2											0.623
PÜ3											0.707
PÜ4											0.780

power for all items and an internal consistency of Cronbach's α between 0.57 and 0.85. However, confirmatory factor analysis uncovered significant weaknesses

in the construct. This suggests that the underlying model only fits unsatisfactory with the empirical data. The following discussion is based on these results.

Table 5: Correlations r between the subscales. Abbreviations see attachment 1.

	SV	CA	FB	TW	PC	PR	WA	AR	FE	RT	PS
SV	*	0.591	0.153	0.427	0.267	0.421	0.547	0.788	0.310	0.154	0.326
CA		*	0.246	0.436	0.210	0.409	0.587	0.752	0.406	0.284	0.489
FB			*	0.033	0.120	0.035	0.097	0.118	0.156	0.339	0.004
TW				*	0.374	0.314	0.531	0.553	0.139	-0.025	0.207
PC					*	0.275	0.218	0.302	0.110	0.203	0.387
PR						*	0.395	0.448	0.264	0.153	0.382
WA							*	0.602	0.183	0.152	0.257
AR								*	0.411	0.192	0.480
FE									*	0.196	0.285
RT										*	0.366
PS											*

Compared to the initial study sample of Boor et al., the present replication study recruited less residents (255 residents vs. 600 residents). Nevertheless, sample size exceeded the lower bound of five residents per item and therefore was regarded as big enough for a factor analysis. In this respect, the replication study has presuppositions comparable to the study of Boor et al. [1]. While the sample of Boor et al. recruited its participants out of 26 specialist areas, residents of the replication study were recruited from the field of anesthesiology, internal medicine, pediatric surgery and family medicine (see Table 2). With respect to item response, only subscale 4. "Teamwork" had comparable distributions, while almost all other scales had lower item values in the replication study (see attachment). For the subscale 3 "Feedback" (items 13 & 14) this may be explained by the fact, that the included hospitals had not established standardized surveillance sheets (i.e. Mini-CEX).

Apart from this scale, low values (<2) were, yet found in item 10 ("My attendings occasionally observe me taking a history.") and in almost all items of subscale 10 "Role of the specialty tutor" (items 42-46). This contrasts the results of Boor et al. who did not find values below 3.0 in the respective items (except for item 10). The study of van Vendeloo et al. framed in the setting orthopedic further education, they also found a global mean of the D-RECT of 3.8 ± 0.4 [9]. In general the dutch samples assessed the global learning climate of further medical education higher than our sample. However, it cannot be ruled out that this mean difference might be due to a specific job-related effect. Thus, it has to be questioned whether specialist training in different disciplines are comparable in principle. Another interpretation simply takes into consideration that such offers are of poorer quality in Germany. Therefore, apart from a possible selection bias a county specific impact on medical training has to be taken into account.

Regarding the internal consistency as a parameter for the reliability of the instrument, the current study shows lower values in five of the subscales, similar values in

two subscales and higher values in four subscales compared to the original work of Boor et al. Apart from the subscales "Supervision", "Feedback", "Peer collaboration" and "Work is adapted to residents' competence", which lay below a Cronbach's alpha of 0.6, all other subscales in our study had a sufficiently high Cronbach's α of at least 0.7 [8]. In particular all subscales with six to eight items showed good coefficients above a value of 0.8, whereas all other subscales failed to pass the lower bound of 0.7 [6].

Discriminatory power in all items was above 0.3. Thus it can be concluded that the D-RECT can distinguish between participants with low (unfavourable appreciation of learning climate) and high scores (good appreciation of learning climate).

Confirmatory factor analysis uncovered fundamental weaknesses in the eleven-factorial model of Boor et al. Analysis of loading weight as well as intercorrelation of the subscales point towards a different underlying structure of the D-RECT-German compared with the original instrument. In particular the scales "Supervision", "Attendings' role" and "Coaching and assessment" showed a high amount of interdependency. With respect to item loadings the scales "Attendings' role" and "Coaching and assessment" should be critically scrutinized on the basis of our results. A not reported explorative factor analysis with oblique rotation supported the inconsistency within the factors. This first of all has to be interpreted that there is insufficient evidence for validity of the international version of the D-RECT questionnaire. Therefore, the use of the D-RECT in the German speaking area should only be considered after stable replication of the results in further studies. In the light of the available evidence an international comparability, as discussed by Boor et al., is not given. Further validation studies with the original questionnaire are absolutely necessary for D-RECT to meet the requirements for a valid evaluation instrument. As a restriction for international comparisons, only the global score of the questionnaire can be used while the

interpretation on the level of subscales is only partially supported by the current evaluation.

Subsequent Dutch studies evaluated external validity of the D-RECT by means of correlation of the D-RECT with other scales: an actual study of Lombarts et al. found correlations between the global score of the D-RECT and the quality of teaching questionnaire SETQ (modified SF-DP26-questionnaire) [10]. Apart from significant positive correlations of the mean scores of both scales, quality of teaching (SETQ) also correlated with the D-RECT subscales 2. "Coaching and assessment", 7. "Work is adapted to residents' competence" and 9. "Formal education". In addition the study of van Vendeloo et al. found, that high global scores of the D-RECT were associated with better quality of life, higher work-life balance, less symptoms of emotional exhaustion and less signs of depersonalisation [9]. Although these results are interesting, they are not sufficiently proved at last. After clarification of the validity of the D-RECT they should be taken into consideration for further research in the field of medical education.

Limitation

Although the current study fulfills the criteria of study quality as provided in Boor et al., the diverging number of specialist areas involved in the studies has to be taken into account. More focused future research with respect to the individual disciplines might safely rule out subject specific differences. Due to the low sample size appropriate multilevel models were not applied in the present study. In both studies a selection-bias cannot be excluded due to voluntary participation of the residents. Finally country specific differences in further education should be taken into consideration, which in the present study design could not be differentiated due to methodological issues.

5. Conclusion

The present study investigated the psychometric properties of the German replication of the D-RECT questionnaire by means of reliability analyses and confirmatory factor analysis. We found structural differences with respect to factorial validity underpinning the need of further validation studies. Although the D-RECT-German could be a helpful tool to evaluate further medical education in the German speaking area the present state of evidence highly demands further studies to examine criterias for test quality. For international comparisons of the two instruments the global score might be used with reservations, until other studies with the D-RECT have been completed.

6. Acknowledgements

The authors thank all residents and attendings who participated in this study, in particular Dr. Klarke Boor, PhD,

Amsterdam, Niederlanden for her support and Professor Dr. med. Martin R. Fischer, LMU Munich, Germany for his important impetus in the practical implementation of this study.

Competing interests

The authors declare that they have no competing interests.

Attachments

Available from

<http://www.egms.de/en/journals/zma/2015-32/zma000997.shtml>

- Attachment 1.pdf (117 KB)
Represented are item and scale characteristics of the items from D-RECT (M: mean; SD: standard deviation, rit: discriminatory power according to Pearson, α =Cronbach's α). The evaluation occurs using a five-point Likert Scale (see paragraph 2.1). #=Reference data (Boor et al., [7]); Items are ordered by the subscales / factors 1. – 11. of the original [7].

References

- Boor K, Van Der Vleuten C, Teunissen P, Scherpbier A, Scheele F. Development and analysis of D-RECT, an instrument measuring residents' learning climate. *Med Teach.* 2011;33(10):820-827. DOI: 10.3109/0142159X.2010.541533
- Diehn T, Niebuhr C. Wettbewerb exzellente Lehre. In: Krull W, Lorentz B, Schlüter A (Hrsg). *Lehre neu denken- Die Zukunft des akademischen Lehrens & Lernens.* Essen: Stifterverband-Verwaltungsgesellschaft für Wissenschaftspflege mbH; 2010.
- Epstein RM. Assessment in medical education. *N Engl J Med.* 2007;356(4):387-396. DOI: 10.1056/NEJMra054784
- Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. *Lancet.* 2001;357(9260):945-949. DOI: 10.1016/S0140-6736(00)04221-5
- Rotem A, Bloomfield L, Southon G. The clinical learning environment. *Isr J Med Sci.* 1996;32(9):705-710.
- Lienert GA, Raatz U. Berechnung von Schwierigkeitsindex, Trennschärfenkoeffizient und Aufgabeninterkorrelation. In: Lienert GA, Raatz U (Hrsg). *Testaufbau und Testanalyse.* Weinheim: Beltz, Psychologie-Verl.-Union; 1994. S.73-113.
- Moosbrugger H, Kelava A. Deskriptiv statistische Evaluation von Items (Itemanalyse) und Testwertverteilungen. In: *Testtheorie und Fragebogenkonstruktion.* Berlin, Heidelberg: Springer-Verlag; 2012. S.75-102.
- Bortz J, Döring N. Hypothesengewinnung und Theoriebildung. In: Bortz JD, Döring N (Hrsg). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler.* Berlin, Heidelberg, New York, Tokio: Springer; 2006. S.355-396. DOI: 10.1007/978-3-540-33306-7_6
- van Vendeloo SN, Brand PL, Verheyen CC. Burnout and quality of life among orthopaedic trainees in a modern educational programme: importance of the learning climate. *Bone Joint J.* 2014;96-B(8):1133-1138. DOI: 10.1302/0301-620X.96B8.33609

10. Lombarts KM, Heineman MJ, Scherpbier AJ, Arah OA. Effect of the learning climate of residency programs on faculty's teaching performance as evaluated by residents. PloS one. 2014;9(1):e86512. DOI: 10.1371/journal.pone.0086512
11. Genn JM. AMEE Medical Education Guide No. 23 (Part 2): Curriculum, environment, climate, quality and change in medical education - a unifying perspective. Med Teach. 2001;23(5):445-454.
12. Roff S, McAleer S. What is educational climate? Med Teach. 2001;23(4):333-334. DOI: 10.1080/01421590120063312
13. Streiner DL. Figuring out factors: the use and misuse of factor analysis. Can J Psych Rev. 1994;39(3):135-140.

Please cite as

Iblher P, Zupanic M, Ostermann T. The Questionnaire D-RECT German: Adaptation and testtheoretical properties of an instrument for Evaluation of the learning climate in medical specialist training. GMS Z Med Ausbild. 2015;32(5):Doc55.
DOI: 10.3205/zma000997, URN: urn:nbn:de:0183-zma0009971

This article is freely available from

<http://www.egms.de/en/journals/zma/2015-32/zma000997.shtml>

Received: 2014-08-17

Revised: 2015-09-15

Accepted: 2015-10-13

Published: 2015-11-16

Copyright

©2015 Iblher et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. See license information at <http://creativecommons.org/licenses/by/4.0/>.

Corresponding author:

Dr. med. Peter Iblher, MME
University of Lübeck, Clinic for Anaesthesiology and
Intensive Care Medicine, Ratzeburger Allee 160, D-23538
Lübeck, Germany, Phone: +49 (0)451/500-2766, Fax:
+49 (0)451/500-3405
peter.iblher@uk-sh.de

Der Fragebogen "D-RECT-German": Adaptation und testtheoretische Güte eines Instruments zur Evaluation der klinischen Weiterbildung

Zusammenfassung

Zielsetzung: In der Arbeitsgruppe von Boor et al. [1] wurde der Fragebogen D-RECT (Dutch Residency Educational Climate Test) zur Erfassung des Lernklimas in der ärztlichen Weiterbildung entwickelt und validiert. Die deutschsprachige Version (D-RECT-German) wird in der vorliegenden Studie testtheoretisch überprüft.

Fragestellung: Ist eine Replikation der Ergebnisse aus der Originalarbeit von Boor et al. als Beleg der Validität des D-RECT möglich?

Methodik: Die Befragung erfolgte onlinebasiert mit dem Fragebogen D-RECT-German. Mit Item- und Reliabilitätsanalysen wurden die Kennwerte der 50 Items in 11 Subskalen sowie die interne Konsistenz (Cronbach's α) ermittelt. Die Validitätsprüfung erfolgte mit einer konfirmatorischen Faktorenanalyse unter Verwendung eines Maximum likelihood basierten Strukturgleichungsmodells.

Ergebnisse: Die Itemanalysen dieser Replikationsstudie mit 255 WBA an 17 deutschen Krankenhäusern ergaben für die Items heterogene Trennschärfen, auch die interne Konsistenz der Subskalen zeigte variable Werte für Cronbach's α zwischen 0.57 und 0.85. In der konfirmatorischen Faktorenanalyse wiesen 6 Items standardisierte Regressionskoeffizienten <0.5 auf die vorgegebenen Dimensionen auf, von denen zwei im Konstrukt „Einstellung der betreuenden Fachärzte“ zu finden waren. Die Korrelationen der Faktoren untereinander wies mit Korrelationen von über 0.7 starke Interdependenzen zwischen den Faktoren „Supervision“, „Coaching“ und „Einstellung der betreuenden Fachärzte“ auf.

Schlussfolgerungen: In der Replikationsstudie mit dem D-RECT-German für den deutschsprachigen Raum zeigten sich strukturelle Unterschiede hinsichtlich der faktoriellen Validität, so dass weitere Validierungsstudien für den internationalen Vergleich notwendig sind.

Schlüsselwörter: D-RECT, Lernklima, ärztliche Weiterbildung, WBA, Validierung, Replikationsstudie

Peter Iblher^{1,2}

M. Zupanic^{2,3}

T. Ostermann^{4,5}

1 Universität zu Lübeck, Klinik für Anästhesiologie, Lübeck, Deutschland

2 Universität Witten/Herdecke, Institut für Didaktik und Bildungsforschung im Gesundheitswesen (IDBG), Witten, Deutschland

3 Universität Witten/Herdecke, Studiendekanat Humanmedizin, Witten, Deutschland

4 Universität Witten/Herdecke, Department für Psychologie, Lehrstuhl für Statistik und Methodenlehre, Witten, Deutschland

5 Universität Witten/Herdecke, Institut für Integrative Medizin, Witten, Deutschland

1. Einleitung

1.1. Einführung

Das Medizinstudium unterliegt in den letzten Jahren einem erheblichen Wandel. Deutliche Anstrengungen wurden unternommen, die Ausbildung zukünftiger Ärztinnen und Ärzte durch medizindidaktische Projekte zu verbessern. Diese Bestrebungen finden ihren Ausdruck auch in der veränderten Wahrnehmung des Stellenwerts einer exzellenten Lehre, die sich in verschiedenen Förderinitiativen manifestiert [2]. Die Erlangung der Approbation als Abschluss dieses Ausbildungsabschnittes stellt die erste Grundlage zur Weiterqualifizierung der Absolventen dar. Es gilt nun ihre erworbenen Fähigkeiten und Kenntnisse

als Ärztin/ Arzt in Weiterbildung (WBA) weiter ausbauen mit dem Ziel, einen spezifischen Facharztstitel zu erwerben. Die Ansprüche hinsichtlich der Überprüfung von klinischer Kompetenz und Lernperformance, die mittlerweile selbstverständlich für die medizinstudentische Ausbildung formuliert werden und im Konsens als Notwendigkeit anerkannt sind [3], [4], wären auch für die ärztliche Weiterbildung zu fordern. Die gründliche und umfassende ärztliche Weiterbildung als Grundlage einer guten Ausbildung wird zu Recht formuliert (http://www.bundes-aerztekammer.de/downloads/20130628-MWBO_V6.pdf, zitiert am 16.04.2015) und sollte überprüft werden, um den einzelnen Ausbildern die Interpretation des Status quo zu ermöglichen, konsekutiv adäquate Veränderungen umzusetzen und diese entsprechend wieder zu überprüfen. Dabei stellt das Lernklima mit seinen verschiedenen Facetten die zentrale Grundlage einer effektiven Ausbil-

dung dar [5]. Wünschenswert wäre hier ein flexibles, etabliertes und verlässliches Messinstrument, das relevante Kompetenzbereiche des Lernklimas abbildet, die Stärken und Schwächen der institutionellen Weiterbildungskonzepte offenlegt und darüber hinaus auch im Rahmen von Lehrforschungsprojekten einsetzbar wäre.

1.2. Der Fragebogen D-RECT zur Evaluation des Lernklimas in der Weiterbildung

Mit Hilfe von qualitativen Untersuchungen zur Etablierung eines optimalen Lernklimas in der ärztlichen Weiterbildung extrahierte die Arbeitsgruppe von Boor et al. drei entscheidende interagierende Domänen:

1. Arbeitsumfeld,
2. Fortbildung und
3. Bedürfnisse der WBA.

Im Rahmen dieses theoretischen Konstruktes wurde ein Messinstrument des Lernklimas in der ärztlichen Weiterbildung auf der Grundlage von elf Kategorien entwickelt (Dutch Residency Educational Climate Test/ D-RECT) (7) (siehe Anhang 1). Die Autoren schlossen, dass die weitere Testung und Validierung des Fragebogens für einen Gebrauch im internationalen Setting wünschenswert wäre. Eine Überprüfung der faktoriellen Validität des finalen Fragebogens sei bisher noch nicht erfolgt. Der Fragebogen wurde zur Publikation aus dem Niederländischen ins Englische übersetzt.

Die vorliegende Arbeit untersucht als Replikationsstudie die Anwendbarkeit des Fragebogens D-RECT im deutschsprachigen Raum mit folgender Fragestellung:

1.3. Fragestellung

1. Testtheoretische Überprüfung einer deutschsprachigen Version des D-RECT (D-RECT-German) zur Überprüfung des Lernklimas mittels Item- und Reliabilitätsanalysen zur Ermittlung der internen Konsistenz (Cronbach's α).
2. Überprüfung der faktoriellen Validität des Fragebogens D-RECT-German mittels konfirmatorischer Faktorenanalyse.

2. Material und Methoden

Die Studie wurde der Ethik-Kommission der Universität Witten/Herdecke vorgestellt, von dort bestanden keinerlei ethische Bedenken gegen die Durchführung. Alle Befragten wurden im Vorwege informiert und hatten die Möglichkeit, die Teilnahme an dieser Studie zu verweigern. Die Einwilligung erfolgte durch konkludentes Handeln.

2.1. Fragebogen D-RECT

Der Fragebogen D-RECT besteht aus 50 Items in 11 Subskalen (siehe Anhang 1). Die Bewertung erfolgt auf einer fünf-stufigen Likert Skala (1=trifft nicht zu – 5=trifft

voll zu). Der englisch publizierte Original-Fragebogen D-RECT wurde durch Muttersprachler zunächst aus dem Englischen ins Deutsche, und dann ins Englische zurück übersetzt. Weiterhin wurden als Kontrollvariablen Alter, Geschlecht, Fachgebiet, Name der Klinik und Weiterbildungsjahr erfragt. Die Befragung erfolgte onlinebasiert (Lime Survey) nach Kliniken und Standorten. Alle erhobenen Daten wurden anonymisiert ausgewertet.

2.2. Testtheoretische Überprüfung des D-RECT-German

Zur Ermittlung der Itemkennwerte wurden Analysen zu Mittelwert (M), Standardabweichung (SD) und Trennschärfe nach Pearson (r_{it}) durchgeführt. Der Trennschärfekoeffizient ist dabei nach Lienert die Korrelation der Aufgabenbeantwortung mit dem Summenwert der Skala und Kennwert dafür, in welchem Ausmaß die Differenzierung der Personen durch das Item mit derjenigen durch die Skala als Ganzes übereinstimmt [6]. Dabei gelten Trennschärfen zwischen 0.4 und 0.7 als gut, von 0.2-0.4 als akzeptabel, von 0.1-0.2 als marginal und Werte unter 0.1 als schlecht [7]. Zur Bestimmung der Skalenkennwerte und -interkorrelationen wurden Mittelwerte (M), Standardabweichungen (SD), Homogenitätskoeffizienten (Cronbach's α) und korrigierte Inter-Skalen-Korrelationen (r , nach Pearson) berechnet. Für Gruppenvergleichstestungen der Reliabilität kann dabei ein Cronbach's α ab 0.7 als ausreichend, ab 0.8 als gut bezeichnet werden [8]. Weiterhin wurden die Voraussetzungen für eine Faktorenanalyse mittels Kaiser-Meyer-Olkin (KMO-) und Bartlett's Test geprüft. Bei gegebenen Voraussetzungen wurde eine konfirmatorische Faktorenanalyse durchgeführt.

Um die Stärke der Beziehungen zwischen den Items und den Konstrukten zu prüfen, wurde das Modell für die vorgegebenen Faktoren mit Hilfe eines Strukturgleichungsmodells simultan geschätzt. Die Modelltestungen wurden mit AMOS 20 und R durchgeführt, wobei die Schätzungen auf der Maximum-Likelihood Methode beruhen. Als Parameter für die Modellvalidität wurden neben dem Chi-Quadrat Wert die für den Skalenfit der konfirmatorischen Faktorenanalyse üblichen inkrementellen Fit-Maße Comparative Fit Index (CFI), Tucker Lewis Index (TLI) und Root Mean Square Error of Approximation (RMSEA) berechnet. Ein signifikanter Chi²-Test deutet dabei auf einen schlechten Modell-Fit hin. Für CFI und TLI weisen Werte >0.90 auf einen guten Modell-Fit hin, während ein RMSEA>0.08 auf einen zu hohen Anteil ungeklärter Varianz hindeutet. Zusammenhänge zwischen den Items und den vorgegebenen Dimensionen wurden durch standardisierte Regressionskoeffizienten r_s beschrieben, wobei Werte <0.5 als nicht ausreichend definiert wurden. Die Korrelationen der Dimensionen untereinander wurden auf Basis der geschätzten Kovarianzmatrix durch Korrelationskoeffizienten bestimmt. Korrelationen >0.7 wiesen dabei auf eine nicht vorhandene Unabhängigkeit der Faktoren untereinander hin. Mit Hilfe der R-Prozedur modindices

wurden abschließend mögliche Fehlspezifikationen des Modells identifiziert.

Aufgrund der geringen Stichprobengröße wurden keine weiteren Analysen wie z. B. eine Mehrebenenanalyse durchgeführt.

3. Ergebnisse

Die Stichprobe setzte sich zusammen aus 255 WBA (weiblich: n=129/50.6%; männlich: n=126/49.4%) an 17 deutschen Krankenhäusern (siehe Tabelle 1) aus vier Fachgebieten (siehe Tabelle 2). Das durchschnittliche Alter der Probanden betrug 32±6 Jahre. Die Aufteilung nach Weiterbildungsjahren ist in Tabelle 3 ersichtlich.

Tabelle 1: Aufteilung der Befragten nach Bundesländern und Standorten (n=2: keine Angabe).

Bundesland	Anzahl Standorte	Anzahl Befragte
Schleswig-Holstein	6	104
Mecklenburg-Vorpommern	3	39
Nordrhein-Westfalen	3	33
Berlin	2	40
Bayern	1	15
Hamburg	2	22
Gesamt	17	253

Tabelle 2: Aufteilung der Befragten nach Fachgebieten

Fachgebiet	Häufigkeit	Prozent
Anästhesie	198	77,6
Innere Medizin	36	14,1
Kinderchirurgie	13	5,1
Allgemeinmedizin	8	3,1

Tabelle 3: Aufteilung der Befragten nach Weiterbildungsjahr

Weiterbildungsjahr	Häufigkeit	Prozent
1.-2. Jahr	93	36,5
3.-4. Jahr	80	29,8
5.-6. Jahr	52	20,4
7.-9. Jahr	23	9,0
>9. Jahr	10	4,0

Die Ergebnisse der Itemanalysen sind in Anhang 1 dargestellt. Die größte Zustimmung mit Mittelwerten über 4.0 fanden Item 3 („Mir ist klar, wen ich um Hilfe bitten muss, falls ich professionelle Unterstützung benötige.“; M: 4.3±0.9), Item 35 („Wenn ich einen Facharzt brauche, kann ich jederzeit einen kontaktieren.“; M: 4.4±0.8) und Item 36 („Wenn ich mich mit einem Facharzt beraten muss, finde ich diese zugänglich.“; M: 4.2±0.8). Den

Items 10 („Meine betreuenden Fachärzte beobachten mich ab und zu bei der Anamnese.“; M: 1.4±0.8), Item 13 („Mein Feedback wird durch Beobachtungsformulare strukturiert.“; M: 1.1±0.6), Item 14 („Beobachtungsformulare werden regelmäßig benutzt, um meine Fortschritte zu bewerten.“; M: 1.1±0.6), Item 44 („Die Beurteilungen dieser Rotation sind hilfreiche Unterhaltungen über meine Leistung.“; M: 1.6±1.4), Item 45 („Meine Zukunftspläne werden während der Beurteilungen besprochen.“; M: 1.7±1.5) und Item 46 („Die Beiträge der verschiedenen betreuenden Fachärzte werden während der Beurteilung zur Kenntnis genommen.“; M: 1.3±1.3) wurde am wenigsten zugestimmt mit Mittelwerten unter 2.0. Die Trennschärfen der Items lagen in keinem Fall unterhalb des kritischen Werts von 0.2, sondern ausnahmslos in einem akzeptablen bis guten Bereich (siehe Anhang 1).

Die Ergebnisse der Skalenanalysen sind in Anhang 1 dargestellt. Die größte Zustimmung der WBA erfolgte in der Subskala „Patientenübergabe“ mit einem Mittelwert von 4.1±0.9. Die geringste Zustimmung wurde in der Subskala „Feedback“ (1.5±0.5) und Subskala „Rolle des Fachbereichs-Ausbilders“ (1.7±1.1) ersichtlich. Es zeigte sich für alle Subskalen eine kritische bis befriedigende interne Konsistenz mit Werten von Cronbach's α zwischen 0.57 und 0.85.

Die Überprüfung der Faktorenstruktur des D-RECT durch die konfirmatorische Faktorenanalyse wies auf einen nur unzureichenden Model-Fit hin. Der Chi-Quadrat Wert war mit 2383,576 hochsignifikant ($p < 0.001$) was für eine ungenügende Modellgüte hinweist. Auch die inkrementellen Fit-Maße CFI und TLI erreichten mit Werten von 0.768 und 0.746 nicht annähernd den Bereich eines guten approximativen Modell-Fits. Nur hinsichtlich des absoluten Modell-Fits kann mit einem RMSEA von 0.068 von einer hinreichend guten Datenpassung durch die postulierte Faktorenstruktur ausgegangen werden.

Die Zusammenhänge zwischen den Items und den vorgegebenen Dimensionen zeigten im gewählten Strukturgleichungsmodell mit Werten zwischen 0.225 und 0.957 eine deutliche Heterogenität, wobei 6 Items standardisierte Regressionskoeffizienten < 0.5 auf die vorgegebenen Dimensionen aufwiesen, von denen zwei im Konstrukt „Einstellung der betreuenden Fachärzte“ zu finden waren. Zufriedenstellende Ladungen konnten für die Dimensionen „Supervision“ (r_s zwischen 0.570 und 0.720), „Teamwork“ (r_s zwischen 0.598 und 0.716), „Professionelle Beziehung“ (r_s zwischen 0.596 und 0.700), „Formale Ausbildung“ (r_s zwischen 0.557 und 0.842), „Rolle des Ausbilders“ (r_s zwischen 0.531 und 0.817) und „Patientenübergabe“ (r_s zwischen 0.596 und 0.780) erzielt werden. Jeweils eine Ladung unter 0.5 wurde in den Dimensionen „Coaching“ (CB7: $r_s = 0.416$), „Feedback“ (FB1: $r_s = 0.225$), „Zusammenarbeit der WBA“ (ZW3: $r_s = 0.441$), und „Fähigkeitsadaptierte Arbeit“ (FW3: $r_s = 0.294$) gefunden. Zwei Ladungen unter 0.5 wurden in der Dimension „Einstellung der betreuenden Fachärzte“ gefunden (EB3: $r_s = 0.470$ und EB8: $r_s = 0.359$) (siehe Tabelle 4). Dieser Befund wird auch durch die Analyse der Fehlspezifikation des Modells bestätigt. Auch hier lädt das Item FB1 auf

Tabelle 4: Ladungen (standardisierte Regressionskoeffizienten r_s der Items auf die Dimensionen)

	SV	CB	FB	TW	ZWB	PBÄ	FWB	EBF	FA	RFA	PÜ
SV1	0.686										
SV2	0.720										
SV3	0.581										
CB1		0.517									
CB2		0.646									
CB3		0.691									
CB4		0.772									
CB5		0.694									
CB6		0.665									
CB7		0.416									
CB8		0.581									
FB1			0.225								
FB2			0.905								
FB3			0.957								
TW1				0.598							
TW2				0.713							
TW3				0.716							
TW4				0.608							
ZW1					0.640						
ZW2					0.698						
ZW3					0.441						
PB1						0.586					
PB2						0.700					
PB3						0.596					
FW1							0.602				
FW2							0.646				
FW3							0.294				
FW4							0.641				
EB1								0.783			
EB2								0.797			
EB3								0.470			
EB4								0.658			
EB5								0.690			
EB6								0.615			
EB7								0.609			
EB8								0.359			
FA1									0.557		
FA2									0.749		
FA3									0.842		
FA4									0.815		
RF1										0.554	
RF2										0.531	
RF3										0.603	
RF4										0.800	
RF5										0.782	
RF6										0.817	
PÜ1											0.596
PÜ2											0.623
PÜ3											0.707
PÜ4											0.780

sieben unterschiedliche Dimensionen und ist für die deutsche Version als kritisch zu betrachten. Die Korrelationsanalyse zwischen den vorgegebenen Dimensionen zeigte mit einer Ausnahme („Teamwork“ vs.

„Rolle Ausbilder“ $r=-0.025$) bei allen Skalen positive Korrelationen, wobei die Subskala „Coaching & Beurteilung“ mit den Skalen „Einstellung der betreuenden Fachärzte“ ($r=0.788$) und „Feedback“ ($r=0.752$) die

Tabelle 5: Korrelationen r der Subskalen untereinander

	SV	CB	FB	TW	ZW	PB	FW	EB	FA	RF	PÜ
SV	*	0.591	0.153	0.427	0.267	0.421	0.547	0.788	0.310	0.154	0.326
CB		*	0.246	0.436	0.210	0.409	0.587	0.752	0.406	0.284	0.489
FB			*	0.033	0.120	0.035	0.097	0.118	0.156	0.339	0.004
TW				*	0.374	0.314	0.531	0.553	0.139	-	0.207
ZW					*	0.275	0.218	0.302	0.110	0.203	0.387
PB						*	0.395	0.448	0.264	0.153	0.382
FW							*	0.602	0.183	0.152	0.257
EB								*	0.411	0.192	0.480
FA									*	0.196	0.285
RF										*	0.366
PÜ											*

höchste gemeinsame Korrelation zeigten. Auch die Skalen „Einstellung der betreuenden Fachärzte“ und „Fähigkeitsadaptierte Arbeit der WBA“ zeigten mit einer Korrelation von $r=0.602$ eine kritische Größe. Alle anderen Skalen lagen mit Werten unter 0.6 in einer akzeptablen Diskriminierung. Die Subskala 11 („Einstellung der betreuenden Fachärzte“) wies dabei mit fast allen anderen Subskalen die jeweils höchste Korrelation auf (siehe Tabelle 5).

4. Diskussion

Die Ergebnisse dieser Replikationsstudie ergaben für alle Items akzeptable bis gute Trennschärfen, die interne Konsistenz der Subskalen zeigte Werte für Cronbach's α zwischen 0.57 und 0.85. Allerdings zeigte die konfirmatorische Faktorenanalyse deutliche Schwächen im Konstrukt auf, die darauf hindeuten, dass das geprüfte Modell eine sehr schlechte Passung zu den Daten hat. Die folgende Interpretation der Ergebnisse sollte daher unter diesem Gesichtspunkt betrachtet werden.

Im Vergleich zu der Stichprobe in der Ursprungsstudie von Boor et al. wurden in dieser Replikationsstudie zwar weniger WBA rekrutiert (255 WBA vs. 600 WBA), es wurden aber auch hier mindestens fünf WBA pro Item eingeschlossen, so dass die Stichprobe als groß genug für eine Faktorenanalyse gelten kann und den Voraussetzungen der Studie von Boor entspricht [1]. Während sich das Stichprobenkollektiv von Boor et al. aus 26 Fachdisziplinen rekrutierte, wurden hier WBA aus den Fächern Anästhesiologie, Innere Medizin, Kinderchirurgie und Allgemeinmedizin eingeschlossen (siehe Tabelle 2). Hinsichtlich der Itemanalysen finden sich bis auf etwa gleiche Werte in Subskala 4. „Teamwork“ nahezu durchgehend niedrigere Werte für die Replikationsstudie (siehe Anhang). Dabei sind die niedrigsten Werte in der Subskala 3. „Feedback“ (Item 13 & 14) durchaus dadurch zu erklären, dass in den eingeschlossenen Krankenhäusern keine standardisierten Beobachtungsformulare (z. B. Mini CEX) etabliert verwendet wurden. Niedrige Werte (<2) finden sich sonst in Item 10 („Meine betreuenden Fachärzte beobachten mich ab und zu bei der Anamnese.“) und nahezu in allen Items der Subskala 10 „Rolle des Fachbereich-Ausbilders“ (Item 42-46). Bei Boor et al. zeigten sich im Gegensatz zu dieser Studie bis auf Item 10 keine Mittelwerte unterhalb von 3.0. Auch in der Studie von

van Vendeloo und Mitarbeitern im Rahmen der orthopädischen Weiterbildung zeigte sich in dem Gesamtergebnis des D-RECT ein Mittelwert von 3.8 ± 0.4 [9]. Generell bewerteten also die niederländischen Stichproben das Lernklima auf Gesamtskalenniveau in ihrer Weiterbildung besser als die deutsche Stichprobe. Dabei muss bedacht werden, dass wegen der Unterschiede im Studienkollektiv ein fachspezifischer Effekt in der Mittelwertbildung nicht auszuschließen ist und zu fragen wäre, ob Weiterbildungen in den unterschiedlichen Disziplinen prinzipiell vergleichbar sind. Trotzdem wäre kritisch zu hinterfragen, ob die Weiterbildungsbedingungen in Deutschland möglicherweise einfach schlechter sind. Generell müsste hier interpretatorisch neben einem möglichen Selection-Bias auch der Einfluss von länderspezifischen Unterschieden in der Weiterbildung bedacht werden.

Hinsichtlich der internen Konsistenz als Kenngröße für die Reliabilität zeigen sich für die vorliegende Replikationsstudie im Vergleich zur Studie von Boor et al. in fünf Subskalen niedrigere Werte, in zwei Subskalen in etwa gleiche Werte und höhere Werte in vier Subskalen. Bis auf die Subskalen „Supervision“, „Feedback“, „Zusammenarbeit der WBA“ und „Fähigkeitsadaptierte Arbeit der WBA“, die knapp unterhalb von 0.6 liegen, erfüllen alle Subskalen der aktuellen Studie die Erfordernisse für Gruppenvergleiche, die ein Cronbach's α von mindestens 0.7 als ausreichend ansehen [8]. Dabei zeigen alle Subskalen mit sechs bis acht Items gute Koeffizienten über 0.8, während alle weiteren Subskalen die Grenze von 0.7 verfehlten [6].

Hinsichtlich der Trennschärfen zeigten nahezu alle Items mindestens Werte über 0.3, so dass postuliert werden kann, dass diese zwischen Teilnehmern mit niedriger, d. h. ungünstiger Bewertung und hoher, d. h. guter Bewertung des Lernklimas trennen.

In der konfirmatorischen Faktorenanalyse wies die elf-faktorielle Analyse nach Boor et al. deutliche Schwächen auf. Die Analyse der Ladungsgewichte sowie der Subskalen untereinander weisen darauf hin, dass der Originalfragebogen und der D-RECT-German zumindest partiell eine andere Struktur aufweisen. Insbesondere die Skalen „Supervision“, „Einstellung der betreuenden Fachärzte“ und „Coaching“ zeigten mit hohen Korrelationen deutliche Interdependenzen auf. Auch in den Itemladungen sind die Skalen „Einstellung der betreuenden Fachärzte“ und „Coaching“ als kritisch zu hinterfragen. Eine hier nicht berichtete explorative schiefwinklige Faktorenanalyse

bestätigte die Inkonsistenz in den Faktoren. Dies wäre zunächst dahingehend zu interpretieren, dass für die internationale Version des Fragebogens D-RECT keine hinreichenden Belege für die Validität des Instruments bestehen. Die Verwendung des Fragebogens im deutschsprachigen Raum wäre daher nur nach stabiler Replikation der Ergebnisse in Folgeuntersuchungen möglich. Eine internationale Vergleichbarkeit, wie von Boor et al. diskutiert, erscheint aufgrund der derzeit vorliegenden Ergebnisse jedoch nicht gegeben. Weitere Validierungsstudien des Originalfragebogens sind dafür unbedingt notwendig, damit der D-RECT den Anforderungen als ein valides Evaluationsinstrument gerecht wird. Einschränkend könnten hier für internationale Vergleiche lediglich die Gesamtwerte des Fragebogens herangezogen werden, ein Vergleich auf Subskalenniveau wäre nur partiell gegeben. Niederländische Folgestudien untersuchten im Sinne einer externen Validierung Zusammenhänge zwischen dem D-RECT und anderen Skalen. So wurden in einer aktuellen niederländischen Studie von Lombarts et al. Korrelationen zwischen dem Fragebogen D-RECT zum Lernklima und dem SETQ-Instrument (modifizierter SFDP26-Fragebogen) zur Lehrqualität betrachtet [10], hier fanden sich signifikante positive Korrelationen zwischen den Gesamtwerten der beiden Skalen, sowie der Lehrqualität und den D-RECT Subskalen 2. „Coaching und Beurteilung“, 7. „Fähigkeitsadaptierte Arbeit der WBA“ und 9. „Formale Ausbildung“. Die Arbeitsgruppe von van Vendeloo und Mitarbeitern zeigte in einem orthopädischen Weiterbildungskollektiv, dass hohe Gesamtwerte des D-RECT assoziiert waren mit besser Lebensqualität, höherer Zufriedenheit hinsichtlich der work-life balance, weniger Symptomen der emotionalen Erschöpfung und weniger Zeichen der Depersonalisation [9]. Diese Ergebnisse sind interessant, letztendlich aber noch nicht hinreichend. Sie könnten aber nach Klärung der Validität des D-RECT Bestandteil weiterer Forschungsfragen im Bereich der ärztlichen Weiterbildung sein.

Limitationen

Auch wenn in der aktuellen Studie die durch Boor definierten Kriterien zum Einschluss einer Mindestzahl an Probanden pro Item eingehalten wurden, so gibt es Unterschiede hinsichtlich der rekrutierten Fachbereiche. Hier wären Untersuchungen zu einzelnen Fachdisziplinen hilfreich, um mögliche fachspezifische Unterschiede sicher auszuschließen. Aufgrund der geringen Stichprobengröße wurde jedoch auf entsprechende Mehrebenenanalysen verzichtet. Grundsätzlich ist ein Selection-Bias in beiden Studien nicht auszuschließen, da die Teilnahme freiwillig war. Auch der Einfluss möglicher länderspezifischer Unterschiede in der Weiterbildung sollte bedacht werden, der im vorliegenden Studiendesign methodisch nicht abzugrenzen ist.

5. Schlussfolgerung

In dieser Studie wurde die deutschsprachige Replikation des originären Fragebogens D-RECT zur Evaluation der ärztlichen Weiterbildung mittels Reliabilitätsanalysen und konfirmatorischer Faktorenanalyse untersucht. Es zeigten sich strukturelle Unterschiede hinsichtlich der faktoriellen Validität, so dass weitere Validierungsstudien notwendig sind. Der D-RECT-German könnte für den deutschsprachigen Raum möglicherweise ein hilfreiches Instrument zur Evaluation der ärztlichen Weiterbildung sein, dies setzt aber zum jetzigen Zeitpunkt noch Folgeuntersuchungen hinsichtlich der Testgütekriterien voraus, beispielsweise auch fachspezifische Replikationsstudien. Für internationale Vergleiche der beiden Fragebögen wären derzeit nur unter Vorbehalt die Gesamtskalenwerte heranzuziehen, bis weitere Validierungsstudien abgeschlossen sind.

6. Danksagung

Die Autoren danken allen Weiterbildungsassistenten und Weiterbildungsbeauftragten, die sich an dieser Studie beteiligt haben, insbesondere Frau Dr. Klarke Boor, Amsterdam, Niederlande für ihre Unterstützung und Herrn Professor Martin R. Fischer, LMU München, Deutschland für wichtige Impulse bei der Umsetzung dieser Forschungsarbeit.

Interessenkonflikt

Die Autoren erklären, dass sie keine Interessenkonflikte im Zusammenhang mit diesem Artikel haben.

Anhänge

Verfügbar unter

<http://www.egms.de/en/journals/zma/2015-32/zma000997.shtml>

1. Anhang 1.pdf (123 KB)

Dargestellt sind die Item- und Skalenkennwerte für die Items des D-RECT, (M: Mittelwert; SD: Standardabweichung, rit: Trennschärfe nach Pearson, α =Cronbach's α). Die Bewertung erfolgt auf einer 5-stufigen Likert Skala (siehe Text). #=Vergleichsdaten von (Boor et al., [7]); Items sind geordnet nach den Subskalen / Faktoren 1. – 11. der Originalarbeit [7].

Literatur

1. Boor K, Van Der Vleuten C, Teunissen P, Scherpbier A, Scheele F. Development and analysis of D-RECT, an instrument measuring residents' learning climate. *Med Teach*. 2011;33(10):820-827. DOI: 10.3109/0142159X.2010.541533

2. Diehn T, Niebuhr C. Wettbewerb exzellente Lehre. In: Krull W, Lorentz B, Schlüter A (Hrsg). Lehre neu denken- Die Zukunft des akademischen Lehrens & Lernens. Essen: Stifterverband-Verwaltungsgesellschaft für Wissenschaftspflege mbH; 2010.
3. Epstein RM. Assessment in medical education. N Engl J Med. 2007;356(4):387-396. DOI: 10.1056/NEJMra054784
4. Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. Lancet. 2001;357(9260):945-949. DOI: 10.1016/S0140-6736(00)04221-5
5. Rotem A, Bloomfield L, Southon G. The clinical learning environment. Isr J Med Sci. 1996;32(9):705-710.
6. Lienert GA, Raatz U. Berechnung von Schwierigkeitsindex, Trennschärfe-Koeffizient und Aufgabeninterkorrelation. In: Lienert GA, Raatz U (Hrsg). Testaufbau und Testanalyse. Weinheim: Beltz, Psychologie-Verl.-Union; 1994. S.73-113.
7. Moosbrugger H, Kelava A. Deskriptiv statistische Evaluation von Items (Itemanalyse) und Testwertverteilungen. In: Testtheorie und Fragebogenkonstruktion. Berlin, Heidelberg: Springer-Verlag; 2012. S.75-102.
8. Bortz J, Döring N. Hypothesengewinnung und Theoriebildung. In: Bortz JD, Döring N (Hrsg). Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. Berlin, Heidelberg, New York, Tokio: Springer; 2006. S.355-396. DOI: 10.1007/978-3-540-33306-7_6
9. van Vendeloo SN, Brand PL, Verheyen CC. Burnout and quality of life among orthopaedic trainees in a modern educational programme: importance of the learning climate. Bone Joint J. 2014;96-B(8):1133-1138. DOI: 10.1302/0301-620X.96B8.33609
10. Lombarts KM, Heineman MJ, Scherpbier AJ, Arah OA. Effect of the learning climate of residency programs on faculty's teaching performance as evaluated by residents. PloS one. 2014;9(1):e86512. DOI: 10.1371/journal.pone.0086512
11. Genn JM. AMEE Medical Education Guide No. 23 (Part 2): Curriculum, environment, climate, quality and change in medical education - a unifying perspective. Med Teach. 2001;23(5):445-454.
12. Roff S, McAleer S. What is educational climate? Med Teach. 2001;23(4):333-334. DOI: 10.1080/01421590120063312
13. Streiner DL. Figuring out factors: the use and misuse of factor analysis. Can J Psych Rev. 1994;39(3):135-140.

Korrespondenzadresse:

Dr. med. Peter Iblher, MME
 Universität zu Lübeck, Klinik für Anästhesiologie ,
 Ratzeburger Allee 160, D-23538 Lübeck, Deutschland,
 Tel.: +49 (0)451/500-2766, Fax: +49 (0)451/500-3405
 peter.iblher@uk-sh.de

Bitte zitieren als

Iblher P, Zupanic M, Ostermann T. The Questionnaire D-RECT German: Adaptation and testtheoretical properties of an instrument for Evaluation of the learning climate in medical specialist training. GMS Z Med Ausbild. 2015;32(5):Doc55.
 DOI: 10.3205/zma000997, URN: urn:nbn:de:0183-zma0009971

Artikel online frei zugänglich unter

<http://www.egms.de/en/journals/zma/2015-32/zma000997.shtml>

Eingereicht: 17.08.2014

Überarbeitet: 15.09.2015

Angenommen: 13.10.2015

Veröffentlicht: 16.11.2015

Copyright

©2015 Iblher et al. Dieser Artikel ist ein Open-Access-Artikel und steht unter den Lizenzbedingungen der Creative Commons Attribution 4.0 License (Namensnennung). Lizenz-Angaben siehe <http://creativecommons.org/licenses/by/4.0/>.