Revisiting the D-RECT tool: Validation of an instrument measuring residents’ learning climate perceptions

MILOU E.W.M. SILKENS1*, ALINA SMIRNOVA1,2*, RENEE E. STALMEIJER2, ONYEBUCHI A. ARAH1,3,4, ALBERT J.J.A. SCHERPBIER2, CEEPS P.M. VAN DER VLEUTEN2 & KIKI M.J.M.H. LOMBARTS1

1University of Amsterdam, The Netherlands, 2Maastricht University, The Netherlands, 3University of California Los Angeles, USA, 4UCLA Center for Health Policy Research, USA

Abstract

Introduction: Credible evaluation of the learning climate requires valid and reliable instruments in order to inform quality improvement activities. Since its initial validation the Dutch Residency Educational Climate Test (D-RECT) has been increasingly used to evaluate the learning climate, yet it has not been tested in its final form and on the actual level of use – the department.

Aim: Our aim was to re-investigate the internal validity and reliability of the D-RECT at the resident and department levels.

Methods: D-RECT evaluations collected during 2012–2013 were included. Internal validity was assessed using exploratory and confirmatory factor analyses. Reliability was assessed using generalizability theory.

Results: In total, 2306 evaluations and 291 departments were included. Exploratory factor analysis showed a 9-factor structure containing 35 items: teamwork, role of specialty tutor, coaching and assessment, formal education, resident peer collaboration, work is adapted to residents’ competence, patient sign-out, educational atmosphere, and accessibility of supervisors. Confirmatory factor analysis indicated acceptable to good fit. Three resident evaluations were needed to assess the overall learning climate reliably and eight residents to assess the subscales.

Conclusion: This study reaffirms the reliability and internal validity of the D-RECT in measuring residency training learning climate. Ongoing evaluation of the instrument remains important.

Introduction

The learning climate in residency has been increasingly recognized as an important contributor to high-quality graduate medical education (GME) (WFME 2003; Weiss et al. 2012, 2013). The learning climate can be conceptualized as residents’ perceptions of the formal and informal aspects of education (Roff & McAleer 2001), including perceptions of the overall atmosphere (Genn 2001) as well as policies, practices, and procedures within the teaching hospital (Lombarts et al. 2014). When residents perceive their learning climate as positive, they are more likely to make more use of their existing knowledge base (Shimizu et al. 2013) and effective learning styles (Delva et al. 2004). Furthermore, a positive learning climate is instrumental in preventing resident burnout (Llera & Durante 2014) and can among others promote career satisfaction and professional identity development in residents (Cross et al. 2006).

Due to the significance of the learning climate for resident education, a key role has been assigned to evaluation and improvement of the learning climate (WFME 2003; Nasca et al. 2010). In order to accurately evaluate the learning climate, literature has stressed the importance of enhancing the validity of instruments used in measuring the learning climate (Soemantri et al. 2010; Colombet-Getz et al. 2014). One instrument that has been increasingly used to measure the learning climate is the Dutch Resident Educational Climate Test (D-RECT). The preliminary D-RECT was validated in 2011 (Boor et al. 2011), which lead to the final 50-item
questionnaire. Pinnock et al. (2013) have investigated the internal consistency and applicability of the D-RECT in an Australian teaching hospital, however, no extensive validation of the final 50-item D-RECT has been performed so far. Furthermore, to our knowledge, the concepts of the D-RECT have not yet been tested in their aggregated form. Literature on climate research regards that an organization’s climate constitutes individual responses aggregated to the unit of analysis (Schneider et al. 2011). Similarly, for the D-RECT, this means that individual responses should be aggregated to the level of the department in order to inform a department’s learning climate. As a result, the identified structure of the D-RECT may not be the optimal structure for evaluating the learning climate of the department.

In light of the increasing importance allotted to the learning climate as well as the need for further validity evidence of learning climate instruments, a reevaluation of the initial psychometric properties of the D-RECT is necessary. Since its introduction, D-RECT has been widely used in post-graduate training in the Netherlands as a part of internal quality improvement efforts, which provides an opportunity for a larger-scale analysis of existing data. Our study investigated the internal validity of the D-RECT as well as its internal consistency and generalizability on the resident and the department levels.

Methods

Setting

The study included departments that provide hospital-based residency training in the Netherlands. During a pre-determined period (usually one month once per year), the residents rotating on the service are invited to fill out the D-RECT. D-RECT measurements are usually repeated every year in order to gain insight into and monitor the strengths and weaknesses of a department’s learning climate and fuel quality improvement initiatives.

Data collection

For this study, we included resident evaluations of the departments’ learning climate using D-RECT between January 2012 and December 2013. If a department was evaluated more than once during the study period, only the most recent evaluation was included.

D-RECT evaluations were completed via an online platform or a paper questionnaire depending on the training program. In case of online evaluations, participants were reminded up to three times by e-mail to participate in the online D-RECT evaluations. For paper-based evaluations, no reminders were provided. Participation in the D-RECT evaluations was anonymous and voluntary for all participants.

The institutional ethical review board of the Academic Medical Center of the University of Amsterdam provided a waiver of ethical approval for this study. Written permission was asked and granted from the departments using the web-based platform and from hospitals using paper questionnaires.

The D-RECT questionnaire

The D-RECT was developed based on qualitative research, expert opinion and a Delphi panel by Boor et al. (2011). The initial questionnaire consisted of 75-items rated on a 5-point Likert scale (1 = totally disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = totally agree). An additional “not applicable” option was also included. Exploratory factor analysis with oblimin rotation based on 1276 resident evaluations revealed 50 items covering 11 constructs describing the learning climate, which were confirmed by a confirmatory factor analysis. The subscale reliability (Cronbach’s α) ranged from 0.64 to 0.85. Generalizability analysis showed that 11 resident evaluations were needed to reliably evaluate all subscales and three for the overall score.

Data analysis

To describe relevant characteristics of the study sample and the D-RECT, descriptive statistics and frequencies were used. Evaluations with more than 50% of the items missing were excluded from the analysis. For evaluations with less than 50% missing, data were assumed to be missing at random and imputed using the expectation–maximization (EM) technique. To be able to differentiate between resident and department level evaluations, departments with only one resident evaluation were excluded from the analysis.

In developing the new structure, first, practical importance of the items was discussed within the research group. Items with a lack of practical importance were excluded from the analysis. One-third of the sample containing resident evaluations was randomly selected for exploratory factor analysis (EFA), which was deemed a sufficient sample (Wetzel 2012). Principal axis factoring was chosen rather than principal component analysis (PCA), because PCA tends to inflate factor loadings and limits the possibility for confirmation by CFA and replication in other samples (Wetzel 2012). Similar to Boor et al. (2011), we choose oblique rotation, since the data showed that factors were correlated. Based on the pattern matrix of the first factor solution, items with factor loadings <0.4 were excluded from further analysis (Hatcher & Stepanski 1994). A second EFA was performed on the remaining items. The Kaiser–Guttman criterion (eigenvalue >1.0) was used to determine the number of factors. The soundness of this model was compared with alternate models based on the scree plot and the amount of explained variance as suggested previously (Schonrock-Adema et al. 2009). Items were assigned to the factor on which the factor loading was highest. The placement of each item was further discussed within the research group.

To assess the fit of the structure obtained by the EFA, a confirmatory factor analysis (CFA) was performed on the remaining two-thirds of the sample containing resident evaluations. In order to compare the fit of the new structure to the original, the CFA was then repeated on the same sample using the original 11-factor structure as identified by Boor et al. (2011). The structure with the best fit was tested on the department level using a CFA. Department level data were attained by calculating the means for each question per department.
The CFA models were estimated by using robust maximum likelihood. The fit of the model was assessed by using the standardized root mean square residual (SRMR), the root mean square error of approximation (RMSEA), the comparative fit index (CFI), and the Tucker–Lewis index (TLI). Cut off values for these fit-indices were pre-determined (SRMR <0.08 for good fit and <0.12 for acceptable fit; RMSEA <0.06 for good fit and <0.10 for acceptable fit; CFI and TLI >0.95 for good fit and >0.90 for acceptable fit) (Brown 2006).

The complete sample of resident evaluations as well as the aggregated sample was used to determine internal consistency, inter-scale correlations, and the item-total correlations. Internal consistency of the subscales was checked by calculating Cronbach’s z for each subscale. Cronbach’s z >0.70 was considered satisfactory (Cronbach 1951). Inter-scale correlations were deemed satisfactory when <0.70. The homogeneity of each scale was assessed by item-total correlations, which should be >0.40 (Arah et al. 2011).

Generalizability analysis was conducted using generalizability theory to determine the optimal number of resident evaluations needed for reliable estimation of the subscale and total scores. We regarded the departments to be the unit of analysis and the number of items as fixed. The resulting design was an unbalanced single-facet nested study with persons (p) nested within departments (d) (p,d) (Blok & Norman 2012). We estimated variance components associated with variance across departments (Sd) and persons nested within departments (Sp,d), the reproducibility coefficient (G), and standard error of measurement (SEM) for varying number of trainees for the mean score and the subscale scores. Similar to Boor et al. (2011), we used SEM <0.26 (1.96 × 0.26 × 2 = 1.0) representing a 1 unit “noise level” on the scale 1–5 as the maximum value for 95% confidence interval interpretation.

The CFA was performed using the lavaan package in R statistical software version 3.1.0 (SAS Inc., Cary, NC). Variance components were estimated using UrGENOVA (SAS Inc., Cary, NC). The remaining analyses were performed with SPSS version 20 (SPSS Inc., Chicago, IL).

**Results**

**Study participants**

Between 2012 and 2013, a total of 2347 D-RECT evaluations were completed, of which nine were excluded based on more than 50% missing data. Thirty-two departments were excluded because they had only one D-RECT evaluation. As a result, 2306 resident evaluations for 291 departments in 48 teaching hospitals including five academic teaching hospitals were included in the study. Seventy-two percent of the residents completed online evaluations, yielding a response rate of 62%. The response rate for the paper-based sample (28%) could not be calculated because the number of invited trainees was not collected. However, based on the literature, we suspect a similar or even higher response rate for the paper-based questionnaires (Yarger et al. 2013). A detailed description of the study population is provided in Table 1.

| Table 1. Characteristics of the study population. |
|-----------------------------------------------|
| **Characteristics**                           | **N (%)** |
| Number of teaching hospitals                 | 48 (100%) |
| Academic                                      | 5 (10%)   |
| Non-academic                                  | 43 (90%)  |
| Number of departments                         | 291 (100%)|
| Surgical                                      | 97 (33%)  |
| Medical                                       | 165 (57%) |
| Auxiliary                                     | 29 (10%)  |
| Number of resident evaluations                | 2306 (100%)|
| Gender of respondents                          |          |
| Male                                          | 899 (39%) |
| Female                                        | 1366 (59%)|
| Missing                                       | 41 (2%)   |
| Year of training                              |          |
| 1                                             | 287 (13%) |
| 2                                             | 309 (13%) |
| 3                                             | 259 (11%) |
| 4                                             | 209 (9%)  |
| 5                                             | 156 (7%)  |
| 6                                             | 94 (4%)   |
| Doctor not in training                        | 292 (13%) |
| Fellow                                        | 46 (2%)   |
| Missing                                       | 654 (28%) |

*Surgical: general surgery, obstetrics and gynecology, orthopedic surgery, urology, otolaryngology, ophthalmology, plastic surgery, oral and maxillofacial surgery, thoracic surgery.

*Medical: internal medicine, pediatrics, neurology, cardiology, emergency medicine, psychiatry, rehabilitation medicine, gastroenterology, pulmonology, anesthesiology, radiology, dermatology, intensive care, geriatrics, pathology, sports medicine, nuclear medicine, medical microbiology, radiotherapy.

*Auxiliary: clinical chemistry, clinical genetics, clinical physics, medical psychology, pharmacy.

**Psychometric properties of D-RECT**

Two items (“Observation forms are used to structure my feedback” and “Observation forms are used periodically to monitor my progress”) were deemed redundant, because observation forms are now used to evaluate all residents and were therefore removed (Directive of the Central College of Medical Specialists 2009). Seven hundred sixty nine (769) evaluations were randomly assigned for the EFA and 1537 for the CFA. The first EFA resulted in 10 factors. Thirteen items were removed based on their factor loading, leaving 35 items for further analysis (Appendix, Table A1, available as Supplementary Material).

The final EFA resulted in a 9-factor structure, which explained 65.5% of the total variance. Overall, items clustered into the following subscales: (1) educational atmosphere, (2) teamwork, (3) role of specialty tutor, (4) coaching and assessment, (5) formal education, (6) resident peer collaboration, (7) work is adapted to residents’ competence, (8) accessibility of supervisors, and (9) patient sign-out (Appendix, Table A2, available as Supplementary Material). Table 2 shows the results of the CFA performed on the resident and department levels as well as the results of the original 11-factor structure.

Cronbach’s z for subscales ranged from 0.71 to 0.86 at the resident level and from 0.80 to 0.91 for the department level (Appendix, Table A2). Corrected item-total correlations ranged from 0.41 to 0.75 at the resident level, and 0.53 to 0.84 at the
department level (Appendix, Table A2). Inter-scale correlations ranged from 0.32 to 0.52 for the resident level, and from 0.37 to 0.66 at the department level. Summary statistics for the D-RECT subscales are provided in the appendix (Appendix, Table A3, available as Supplementary Material).

Table A2 reports the variance components and minimum number of trainees needed to reliably assess each subscale and the mean score. The minimum number of resident evaluations needed ranged from three for the mean score to eight for patient sign-out and resident peer collaboration subscales.

### Discussion

#### Main findings

The aim of this study was to test the internal validity and the reliability of the D-RECT on both the resident and the department level. The results showed that the learning climate can be evaluated on both the level of the resident and the department using 35 questions grouped into nine subscales: educational atmosphere, teamwork, role of specialty tutor, coaching and assessment, formal education, resident peer collaboration, work is adapted to residents’ competence, accessibility of supervisors, and patient sign-out. Furthermore, eight residents per department were needed to evaluate all subscales of the learning climate reliably.

#### Explanation of findings

Overall the new structure reflects the original D-RECT questionnaire. Although the subscale “feedback” was dropped, the topic is still represented in the questionnaire by the subscale “coaching and assessment”. Items from the subscale “attendings’ role” were divided into the new scales “educational atmosphere” and “accessibility of supervisors”. The literature also regards clinical teachers as being primarily responsible for creating an atmosphere, in which learners can comfortably identify and address their limitations (Ramani & Leinster 2008), which is represented in the questionnaire by the factor “educational atmosphere”. The subscale “supervision” has been incorporated in the new subscale “accessibility of supervisors”. Similarly, accessibility of timely and appropriate supervision has been regarded to be an important factor in both educational and patient outcomes (Farnan et al. 2012).

In conclusion, we believe that the two new subscales are theoretically representative of the contributing factors to the learning climate in residency.

The new 9-factor structure was supported by the CFA. The SRMR and the RMSEA, indicating how well the a priori model reproduces the sample data, showed a good fit. In contrast, the CFI and TLI, measuring the improvement of our model when compared to a restricted model, showed a slightly lower albeit acceptable performance, especially at the department level. Nevertheless, rather than evaluating single fit indices to accept or reject a model, it has been recommended to consider the fit indices in a combined manner instead (Hu & Bentler 1999). As such, the overall good fit of the new 9-factor model on the resident level has been demonstrated by the acceptable values of the incremental fit indices (CFI and TLI) and the good fit of the absolute fit indices (SRMR and RMSEA). When compared with the fit of the original 11-factor structure in our sample, the 9-factor model showed an improvement in the CFI and TLI indices. At the department level, the incremental fit indices were slightly below the acceptable cutoff points, while the absolute indices indicate good fit. It can be concluded that the 9-factor model showed an acceptable fit at the department level. The applicability of the 9-factor structure at the resident and department levels was further supported by the item-total correlations >0.30 that indicated that each item contributed to the measurement of the concept learning climate, and inter-scale correlations <0.70 that indicated that D-RECT comprised nine sub-constructs.

With regard to the reliability, internal consistency (Cronbach’s α) was satisfactory at the resident as well as the department level. Generalizability analysis, where residents were nested within departments, showed that a minimum of eight resident evaluations was required for the 9-factor questionnaire, whereas previously 11 resident evaluations were needed. The overall climate could still be reliably evaluated with only three residents. Between-resident differences accounted for two to three times more variation in scores than between-department differences.

In choosing our statistical approach, we aimed to retain a number of factors that would maximize the instrument’s explanatory power. We considered this especially important since those that use D-RECT have appreciated its multidimensionality when used in quality improvement activities (Pinnock et al. 2013). In line with this reasoning, the two-item scale (patient sign-out) was retained to support patient safety and residency training (Myers & Bellini 2012).

With regard to the content of the subscales, the constructs of the D-RECT fit within climate frameworks (Schonrock-Adema et al. 2012). Ostroff (1993) have researched the broad concept of climate and have organized climate perceptions into three higher order facets: the affective, the cognitive and the instrumental facet. In the D-RECT, the affective facet is accounted for by the overall feeling of the atmosphere (constructive educational atmosphere), how well residents work together (resident peer collaboration) and how well the inter-professional team works together (teamwork). The cognitive facet is accounted for by the focus on how the supervisor helps the resident reflect on performance (coaching and assessment), to what extent the resident is involved in the

| Table 2. Fit indices of the 9-factor structure at the resident and department level compared with the original 11-factor structure at resident level. |
|-----------------|-----------------|-----------------|
|                  | Resident level  | Department level |
|                  | (n = 1537)      | (n = 291)       |
|                  | 35-item (9-factor) questionnaire | 50-item (11-factor) questionnaire | 35-item (9-factor) questionnaire |
| CFI              | 0.92            | 0.86            | 0.89            |
| TLI              | 0.91            | 0.85            | 0.88            |
| SRMR             | 0.04            | 0.05            | 0.06            |
| RMSEA            | 0.04            | 0.05            | 0.06            |

In conclusion, we believe that the two new subscales are theoretically representative of the contributing factors to the learning climate in residency.
patient hand-over (patient sign-out) and to what extent the work of the resident is adapted to the level of experience of the resident (work is adapted to residents’ competence). The D-RECT takes the instrumental facet into account by evaluating planned education (formal education), what the involvement of the formal educator is (role of the specialty tutor) and to what extent the supervisors are involved (accessibility of supervision). Furthermore, the D-RECT has a more applied grounding. Since the D-RECT is an instrument to measure the learning climate in work-based GME, it is intuitive that the instrument does not only focus on educational activities that are linked to GME, but also to the patient-care related aspects.

**Strengths and limitations of the study**

In addition to contributing to the validity evidence of the D-RECT, this study adds to the literature by using the unit of analysis – the department. It would have been preferable to perform an EFA on the department level to identify the best fitting structure separately for this level. However, the aggregated dataset with only 291 departments was rather small to perform both an EFA and a CFA (Wetzel 2012). Since the number of resident evaluations was much higher compared to the number of departments in our sample, we, therefore, chose an alternative route by first exploring the structure of the D-RECT on the resident level and then confirming it on the department level. A larger sample is more likely to generate factor loadings that closely reflect the population and are therefore less variable in repeated testing (MacCallum et al. 1999). In our study, the EFA is based on a sample exceeding the 10:1 ratio, thereby contributing to the stability of the findings (Wetzel 2012). The number of departments was considered to be sufficient for a CFA, especially since every department score is composited from multiple resident evaluations.

Furthermore, the multicenter setting, with both academic and non-academic hospitals covering various specialties across the Netherlands, affirms the soundness of the results in the Netherlands. However, since learning climate reflects the perceptions of the residents, cultural differences in residents’ expectations may occur (Wong 2011). As a result, residents from other cultures may emphasize aspects of training that are not covered by the D-RECT.

**Implications for practice and future research**

For those who use the D-RECT we hope to bring trust in the structure of the instrument, whilst providing slight nuances in the arrangement of subscales. A significant improvement for practice is the rigorous reduction in the number of items, which may improve response rate and truthfulness of residents’ responses (Colbert-Getz et al. 2014). Another improvement is that D-RECT requires eight instead of 11 residents for evaluation of all scales. This may encourage smaller departments to make use of the D-RECT in the future. In departments with fewer than eight but at least three residents, the overall learning climate can still be reliably evaluated.

While this study furthers the evidence of D-RECT’s internal structure, future research could continue to provide validity evidence, including the response process and relationships with other variables. Specifically, attention should be paid to the formulation of the items, including use of double negatives. Reformulation of such items might improve the responses of the residents. Finally, as already mentioned by Boor et al. (2011), investigating the effects of the D-RECT on practice would be a useful step forward in contributing to the discussion on how to improve the learning climate as part of quality assurance programs.

**Conclusion**

In conclusion, after analyzing the reliability and internal validity of the D-RECT we propose an updated structure, which holds on both resident and department levels and can be used to evaluate the learning climate of departments provided a minimum number of residents per department. With 35 items divided into nine subscales, the instrument is now shorter and may therefore be more suitable for practice and research. However, ongoing evaluation of the applicability of the D-RECT items to practice is needed.

**Glossary**

Educational Climate: The way in which students perceive their educational environment and teaching practices.

Genn JM. 2001. AMEE Medical Education Guide No. 23 (Part 2): Curriculum, environment, climate, quality and change in medical education – A unifying perspective. Med Teach 23:445–454.

**Notes on contributors**

MILOU SILKENS, MSc, is a PhD student at the Professional Performance Research Group, University of Amsterdam, Academic Medical Centre, the Netherlands

ALINA SMIRNOVA, MD, is a PhD student at Maastricht University, the Netherlands

RENEE STALMEIJER, MSc, PhD, is an Assistant professor and Director of the Master of Health Professions Education at Maastricht University, the Netherlands

ONYEBUCHI ARAH, MD, PhD, is a Professor at the Department of Epidemiology, UCLA Fielding School of Public Health, United States of America

ALBERT SCERPBIER, MD, PhD, is a Professor and Dean of the Faculty of Health, Medicine and Life Sciences, Maastricht University, the Netherlands

CEES VAN DER VLEUTEN, MSc, PhD, is a Professor and Scientific director of the School of Health Professions Education at Maastricht University, the Netherlands

KIKI LOMBARTS, MSc, PhD, is a Professor and head of the Professional Performance Research Group, Academic Medical Centre, University of Amsterdam, the Netherlands

**Acknowledgements**

The authors thank their colleagues for their input, Ron Hoogenboom for help with generalizability theory, and trainees who filled out the D-RECT questionnaire.
Declaration of interest: The authors report that they have no conflicts of interest. This research was made financially possible by the Dutch Ministry of Health (VWS) through the project “Quality of clinical teachers and residency training programs”.

References

Arah OA, Hoekstra JB, Bos AP, Lombarts KM. 2011. New tools for systematic evaluation of teaching qualities of medical faculty: Results of an ongoing multi-center survey. PLoS One 6(10):e25983.

Bloch R, van der Vleuten C, Teunissen P, Scherpbier A, Scheele F. 2011. Development and analysis of D-RECT, an instrument measuring residents’ learning climate. Med Teach 33(10):820–827.

Brown TA. 2006. Confirmatory factor analysis for applied research. New York: Guilford.

Cronbach LJ. 1951. Coefficient alpha and the internal structure of tests. Psychometrika 16:297–334.

Duret L, Dronkers JF, Cottencin O, Arah OA. 2013. Characteristics in response rates for surveys administered to Canadian residency programme. Med Educ 47(12):1209–1219.

Finnel T, Wetzel AP. 2012. Factor analysis methods and validity evidence: A review of instrument development across the medical education continuum. Med Teach 34(11):960–992.

Genn JM. 2001. AMEE Medical Education Guide No. 68. Med Teach 23(4):333–334.

MacCallum RC, Widaman KF, Zhang S, Hong S. 1999. Sample size in factor analysis. Psychol Methods 4(1):84–99.

Myers JS, Bellini LM. 2012. Resident handoffs: Appreciating them as a critical competency. J Gen Intern Med 27(3):270–272.

Ostroff C. 1993. The effects of climate and personal influences on individual behavior and attitudes in organizations. Organ Behav Human Decision Process 56(1):56–90.

Pinnock R, Welch P, Taylor-Evans H, Quirk F. 2013. Using the DRECT to assess the intern learning environment in Australia. Med Teach 35(8):699.

Ramani S, Leinster S. 2008. AMEE Guide no. 34. Teaching in the clinical environment. Med Teach 30(4):347–364.

Roff S, McAleer S. 2001. What is educational climate? Med Teach 23(4):335–334.

Schneider B, Ehrrart MG, Macey WH. 2011. Perspectives on organizational climate and culture APA handbook of industrial and organizational psychology, Vol 1: Building and developing the organization. Washington, DC: American Psychological Association. pp 573–594.

Soemantri D, Herrera C, Riquelme A. 2010. Measuring the educational environment in health professions studies: A systematic review. Med Teach 32(12):947–952.

Supplementary material available online

The supplementary material for this paper is available online at http://dx.doi.org/10.3109/0142159X.2015.1060300.