Medical Teacher

ISSN: 0142-159X (Print) 1466-187X (Online) Journal homepage: http://www.tandfonline.com/loi/imte20

Content validity of workplace-based portfolios: A multi-centre study

Nele R.M. Michels, Marijke Avonts, Griet Peeraer, Kris Ulenaers, Luc F. Van Gaal, Leo L. Bossaert, Erik W. Driessen, Arno M.M. Muijtjens & Benedicte Y. De Winter

To cite this article: Nele R.M. Michels, Marijke Avonts, Griet Peeraer, Kris Ulenaers, Luc F. Van Gaal, Leo L. Bossaert, Erik W. Driessen, Arno M.M. Muijtjens & Benedicte Y. De Winter (2016) Content validity of workplace-based portfolios: A multi-centre study, Medical Teacher, 38:9, 936-945, DOI: 10.3109/0142159X.2015.1132407

To link to this article: http://dx.doi.org/10.3109/0142159X.2015.1132407

2016, 38: 936–945 MEDICAL TEACHER

Content validity of workplace-based portfolios: A multi-centre study

NELE R.M. MICHELS[1], MARIJKE AVONTS[1], GRIET PEERAER[1], KRIS ULENAERS[1], LUC F. VAN GAAL[1], LEO L. BOSSAERT[1], ERIK W. DRIESSEN[2], ARNO M.M. MUIJTJENS[2] & BENEDICTE Y. DE WINTER[1]

[1]University of Antwerp, Belgium, [2]University of Maastricht, the Netherlands

Abstract

Background: Portfolios are used as tools to coach and assess students in the workplace. This study sought to evaluate the content validity of portfolios as reflected in their capacity to adequately assess achieved competences of medical students during clerkships.

Methods: We reviewed 120 workplace portfolios at three medical universities (Belgium and the Netherlands). To validate their content, we developed a Validity Inventory for Portfolio Assessment (VIPA) based on the CanMEDS roles. Two raters evaluated each portfolio and indicated for each VIPA item whether the portfolio provided sufficient information to enable satisfactory assessment of the item. We ran a descriptive analysis on the validation data and computed Cohen's Kappa to investigate interrater agreement.

Results: The portfolios adequately covered the items pertaining to the communicator (90%) and professional (87%) roles. Coverage of the medical expert, collaborator, scholar and manager roles ranged between 75% and 85%. The health advocate role, covering 59%, was clearly less well represented. This role also exhibited little interrater agreement (Kappa < 0.4).

Conclusions: This study lends further credence to the evidence that portfolios can indeed adequately assess the different CanMEDS roles during clerkships, the health advocate role, which was less well represented in the portfolio content, excepted.

Background

Portfolios are used as coaching and assessment instruments to support and evaluate the learning and performance of students in the workplace (Miller 1990; Royal College of General Practitioners 1993; Snadden & Thomas 1998; Friedman Ben David et al. 2001; Davis et al. 2001; McCready 2007; Driessen et al. 2007; Colbert et al. 2008). The literature abounds with definitions of 'portfolio' used for educational purposes. Buckley et al. (2009) framed it as 'a collection of evidence of student activity that outlines the student's own learning experience, requires some 'intellectual processing' on the part of the student, *and* draws together more than one item' (Buckley et al. 2009). It has been demonstrated that portfolios help to improve students' reflective ability, their personal and professional development, and their communication and critical thinking skills (Driessen et al. 2006; McCready 2007; Driessen et al. 2007). Measuring the quality of portfolio assessment is important but can be daunting. Whereas the reliability of portfolio assessment has been studied intensively, there is little concordance between outcomes (Pitts et al. 2001; Driessen et al. 2007; Michels et al. 2009). By contrast, there is a paucity of appropriate research into the *validity* of portfolio assessment (Driessen et al. 2007; Buckley et al. 2009; Tochel et al. 2009).

Practice points

- A portfolio can be used as an effective instrument to evaluate students' achievements in clerkships. More specifically, a whole gamut of competences can be assessed using particular portfolio elements.
- To enhance content validity of portfolio assessment, we need clearly determined aims and blueprints, in addition to an appropriate selection of assessment instruments fitting the intended competences.
- This study calls on curricula and portfolio designers to explore new portfolio tools that enable adequate assessment of specific competences that currently lack sufficient coverage.

Though present-day approaches to validity research are manifold, the general approach is to investigate validity through an ongoing process of proving or disproving the correlation between the purpose of an instrument and the final construct (Messick 1995; Kane 2001; Downing 2003; Kane 2006a, b). It is also generally agreed that validity encapsulates multiple dimensions each of which should be investigated. The Standards of Educational and Psychological Measurement define five such dimensions, referred to as sources of validity,

Correspondence: Nele R.M. Michels, University of Antwerp, Campus Drie Eiken, R3.14, Universiteitsplein 1, 2610 Antwerp (Wilrijk), Belgium. Tel: +0032 3 265 29 35; Fax: +0032 3 265 25 26. E-mail: nele.michels@uantwerpen.be

ISSN 0142-159X print/ISSN 1466-187X online/16/090936–10 © 2016 Taylor & Francis
DOI: 10.3109/0142159X.2015.1132407

being: content, response process, internal structure, relationship to other variables and consequences (Downing 2003). In a former study we investigated one of these sources, that is, the internal structure of a portfolio assessment, by performing a generalisability analysis (Michels et al. 2009). The present study, however, is concerned with content validity, which can be defined as evidence of content relevance and representativeness of the test construct. Hence, content validation provides information about the data obtained from an assessment instrument and the inferences that can be drawn from those data (Haynes et al. 1995). Since portfolios are often used as high-stakes evaluations, we consider content validity as an important aspect of validity. In the current validity study we have sought to investigate whether and to what extent the competences a portfolio is supposed to measure are, indeed, reflected in the content of students' high-stakes portfolios.

Methods

We validated the content of workplace-based clerkship portfolios at three different medical schools, specifically the University of Antwerp (Belgium), Maastricht University (the Netherlands), and Utrecht University (the Netherlands). To collect appropriate and sufficient data, we constructed a new inventory based on the CanMEDS competencies. As this will enhance the interpretation of the study results, we will first present the purposes, contexts, and procedures of the selected portfolios in a separate 'setting' section before describing the development of the inventory and the set-up of the validity study.

Setting

We compiled a comprehensive register encompassing essential contextual information on the three portfolios involved (Table 1). To that end, we consulted faculty staff, perused the portfolio guidelines and explored the literature on the three institutions' portfolios (Michels et al. 2009, 2010; Driessen et al. 2012).

The portfolios of the medical schools of Antwerp (Michels et al. 2009, 2010), Maastricht (Driessen et al. 2012) and Utrecht under scrutiny all formed a compulsory part of undergraduate clerkships that required students' full-time involvement in a single (Maastricht) or a number of (Antwerp and Utrecht) disciplines. Whilst the portfolios of Antwerp and Maastricht related purely to clinical clerkships, those of Utrecht could also relate to scientific and educational clerkships. The portfolios served either one of two purposes, or both, specifically to assess whether students had achieved the intended competences during the clerkship (portfolio as assessment instrument), and/or to guide the students in their mastery of these competences (portfolio as coaching instrument). The medical school of Utrecht wielded the portfolio for assessment purposes only, whereas the medical schools of Maastricht and Antwerp extended its use to include coaching as well. The main emphasis of the latter university, however, remained on assessment. In both universities students received personal coaching from tutors, who, in the case of Antwerp, were skills lab teachers, and, in the Maastricht situation, were clerkship mentors (i.e. clinical supervisors in the workplace). At the end of the assessment procedure, all universities produced feedback forms listing the assessment criteria and required competences together with students' final ratings. Considering the relevant feedback these forms contained, we regarded them as coaching instruments too.

Slight differences between the three institutions also existed with respect to the assessment procedure. Utrecht medical school, for instance, combined written and oral assessments: at the end of each clerkship the local supervisor assigned a rating (written assessment), which was then discussed, together with the overall portfolio and other ratings, with one of the assessors of a fixed assessment team who decided on a pass or fail (final oral assessment). A double written assessment procedure was in place at the University of Antwerp where the final score was made up of assessments by skills lab teachers and clinicians (Michels et al. 2009, 2010). Maastricht University, conversely, followed an assessment procedure based on triangulation, meaning that three types of assessment formed the input for an assessment meeting held between mentor and student: (1) a portfolio rating ranging from insufficient to excellent assigned by an assessment committee; (2) an assessment form filled in by the mentor; and (3) a self-assessment form filled in by the student (Driessen et al. 2012). Both Maastricht and Utrecht medical schools organised intermediate assessments to monitor progress and, if needed, to remediate.

Since the intent of all portfolios under scrutiny was to measure and/or reflect upon students' competences, they were all—eventually at least—largely premised on the constructs of the CanMEDS Roles Framework. Utrecht University was the first to take this approach in 2005, followed by Maastricht University a year later. The University of Antwerp took a detour by first introducing the clerkship portfolio in 2004 based on principles of the three-circle model (The Scottish Doctor) (Harden et al. 1999; Simpson et al. 2002), and later transitioning to the CanMEDS Roles Framework (CanMEDS 2000; Frank 2005). The last column of Table 1 presents the elements of which the institutions' portfolios were composed, listing a host of instruments or assignments, from written assessments and workplace-based assessments to self-reflections, self-assessments and mentoring or assessment reports. Basically, in Antwerp, students were expected to make reports on patients seen, give presentations and write self-reflections and a personal development plan (PDP). The Maastricht portfolio, by contrast, contained not only self-reflections and a PDP, but also direct workplace-based assessments or observations, such as mini Clinical Evaluation Exercises (mini-CEXs), Multi-Source Feedback (MSF) assessments and CATs (Critical Appraisals of a Topic). Utrecht, finally, used a combination of elements included in the Antwerp and Maastricht portfolios, complemented by self-assessments, Professional Behaviour (PB) assessments, clerkship observations, clerkship assessments and peer reviews.

Development of the inventory

In a previous study, we had developed an inventory that measured the presence/absence of CanMEDS competencies. By means of a three-round Delphi method, we invited experts

Table 1. Description of the context, purpose, coaching and assessment of the portfolios reviewed.

	1) Context			2) Purpose	3) Coaching					
					Tutorship					
	Related to	Time period	Mandatory		Mandatory	Type	Tutor	Feedback (after assessment)		
University of Antwerp	Clerkship (clinical), UG, year 6 hospital (different disciplines) or GP	8 months	Yes	+/− coaching portfolio assessment portfolio	No	Individual, ≥1 tutor contacts	Teacher skills lab	Feedback form (after final assessment)		
Maastricht University	Clerkship (clinical), UG, year 6 hospital, GP: 1 optional discipline	4 months	Yes	Coaching portfolio Assessment portfolio	Yes	Individual, 3 tutor contacts	Clerkship mentor (clinician)	Feedback form (after final assessment)		
Utrecht University	Clerkship (clinical, scientific, educational) UG, year 6 hospital (different disciplines), GP, nursing home, social medicine	11 months	Yes	Assessment portfolio	No tutorship unless identified problems			Feedback form (after clerkship assessments and final assessment)		
	4) Assessment							5) Portfolio elements		
	Procedure									
	Assessment type	Rating	Rater	Intermediate	Score	Contribution	Assessment's aim			
University of Antwerp	Written assessment	Double rating	Skills lab teacher + clinician	No	Mark	Moderate to high-stakes	Assessing competences (TCM & CanMEDS)	Case report (*N* = 4) Self-reflection (*N* = 3) Presentation (*N* = 2) Obstetric report (*N* = 1) Surgery report (*N* = 1) Financial analysis (*N* = 1)	PDP (*N* = 1) GP task (*N* = 1) Profiling space Autopsy (*N* = 2) Self-care reflection (*N* = 1)	
Maastricht University	Written assessment	Single rating + triangulation	Assessment committee	Yes (mentor)	Insufficient/sufficient/good/excellent	High stakes	Assessing competences (CanMEDS)	PDP Mini-CEX (*N* = 5) MSF (*N* = 2) CAT (*N* = 2)	Mentoring report Reflections Profiling space	
Utrecht University	Written assessment per clerkship	Double rating	Clinician (clerkship)	Yes	Mark per clerkship	High stakes	Assessing competences (CanMEDS)	Case report FACR report EBC report Self-reflection Self-assessments PB assessment	Presentation Deepening task Article/publication Progression report Peer review Referral letter	Mini-CEX Observations Clerkship assessment CD PDP
	oral assessment at the end	single rating	fixed assessment team		pass & fail					

Register of the most important features of the portfolios reviewed. We distinguished 1) portfolio context, 2) purpose of the portfolio implementation, 3) the portfolio coaching, 4) the portfolio assessment and 5) the portfolio elements. In this last column *N* is provided between brackets when a fixed number is required.

CAT, critical appraisal of a topic; CD, competence development; EBC, evidence-based case; FACR, follow analyse clarify reflect; GP, general practice; mini-CEX, mini clinical evaluation exercise; MSF, multisource feedback; PB, professional behaviour; PDP, personal development plan; TCM, the Three Circle model; UG, undergraduate.

Table 2. The Validity Inventory for Portfolio Assessment (VIPA).

Roles and items	+	+/−	−
Medical expert			
Medical knowledge (1 a, 5)*			
Acquired skills (2)			
Medical decision-making (1 b)			
(Multidisciplinary) health care plan (3, 4)			
Integration of the CanMEDS roles (6)			
Communicator			
Intake and history taking (7)			
Communication in the doctor–patient(-family) relationship (12, 13)			
Communication of scientific research (11)			
Communication of patient cases (verbal and written) (8–10)			
Reflection on own communication (14)			
Collaborator			
Appropriate contribution in a team of health care providers (15–17)			
Teamwork tasks (18)			
Reflection on team participation (19)			
Manager			
Reflection on self-care and balance between work-personal development (20)			
Priorities, administrative and organisational tasks, information technology (21–24)			
Health care system, insight into costs and contractual procedures (25–27)			
Health advocate			
Reflection on aspects influencing general health (bio, psychosocial and existential) (28)			
Primary and secondary prevention, patient safety (29–31)			
Patient's benefits, guidance in health care (32–34)			
Reflection on critical incidents (35)			
Scholar			
Scientific thinking and acting (36–39)			
Attitude of lifelong learning—a personal development plan (40, 44)			
Transmission of medical knowledge and practice to the scientific community and an evolving society (41–43)			
Professional			
Highest quality of care with an appropriate attitude and behaviour (45, 46, 48)			
Reflection on professional behaviour and attitude (47, 49)			

The VIPA is an inventory we developed on the basis of a previous inventory [CCBI (Michels et al. 2012)]. It contains the seven CanMEDS Roles, each of which consists of 2–5 items. The numbers reported after each item (*) refer to the competencies of the original CCBI (see Appendix).

When evaluating a portfolio, a rater selects one of three codes for each item: the portfolio provided sufficient (+), moderate (+/−), or insufficient (−) information to enable satisfactory assessment of the specific item.

with an educational and medical background (N= 25–30) to indicate for each CanMEDS role (on a 6-point Likert scale) the extent to which the individual competencies were of relevance to workplace-based portfolios. In addition to this, we asked the experts to help us redefine and reformulate the competencies to suit the specific context of workplace-based portfolios. We ran a descriptive analysis on the data and feedback received, leading to the final CanMEDS Competency-Based Inventory (CCBI) (Michels et al. 2012) (see Appendix 1). To be able to evaluate the content validity of portfolios for the present study, we modified the CCBI to be more compact and hence more practical. We did so by conducting semi-structured, audiotaped interviews with nine randomly selected experts (via www.random.org) from the previous Delphi expert pool (Michels et al. 2012), during which we discussed a draft version of the shortened inventory. The transcription and thematic analysis of the interviews (Braun & Clarke 2006) resulted in the new Validity Inventory for Portfolio Assessment (VIPA) that consisted of 25 items, with 2–5 items describing each role (see Table 2).

Validity study

For this study we randomly selected 40 portfolios per medical school during the period spanning 2011–2012. Two raters independently rated each portfolio using the VIPA (Table 2). In total there were three raters (MA, GP and KU) who formed three rater pairs. Although all raters were affiliated to the University of Antwerp, they had been outsiders to the development and evaluation process of the Antwerp portfolio. For each portfolio, raters indicated per VIPA item whether the portfolio provided sufficient information to enable adequate assessment of that item. The answer options were three: (1) the information provided was sufficient, coded as '+'; (2) the portfolio did provide information, but only moderately, coded as '+/−'; and (3) the portfolio provided no or too little information to enable assessment of the item, coded as '−' (see Table 2). Before the raters started the review process, they studied the CCBI and VIPA and performed a trial rating on 2 or 3 portfolios.

Upon conclusion of the review process, we counted frequencies of the VIPA ratings ('+', '+/−' or '−') for each item *and* for each role and converted these to percentages of the total number of ratings per item/role. These percentages reflected the extent to which the portfolio provided sufficient information to enable assessment of the individual VIPA items. In addition, we investigated the VIPA's interrater agreement, by computing Cohen's Kappa for each item and rater pair. In doing so, we summarised Kappa per item by calculating the

Table 3. Frequencies of workplace-based portfolio ratings per item, presented as percentages of the total number of ratings per item/role; Percentage of Agreement and Kappa for rater pairs as regards the 3-point classification of VIPA items.

	Frequencies ($N = 120$)			Interrater Agreement	
Roles & items	+	+/−	−	PA	KAPPA
Medical expert					
Medical knowledge	100			100	–
Acquired skills	89.2	10.8		97	0.80
Medical decision-making	99.6	0.4		99	–
(Multidisciplinary) health care plan	98.75	1.25		98	–
Integration of the CanMEDS roles	22.5	77	0.4	84	0.62
Total percentage	**82.0**				
Communicator					
Intake and history taking	100			100	–
Communication in the doctor–patient(-family) relationship	82	17.5	0.4	83	0.39
Communication of scientific research	100			100	–
Communication of patient cases (verbal and written)	72	28		90	0.78
Reflection on own communication	97	3		96	0.29
Total percentage	**90.2**				
Collaborator					
Appropriate contribution in a team of health care providers	71.2	28	0.8	93	0.86
Teamwork tasks	81.3	16.7	2	84	0.48
Reflection on team participation	93	6.2	0.8	91	0.33
Total percentage	**81.8**				
Manager					
Reflection on self-care and balance between work-personal development	77.5	11	11.5	83	0.49
Priorities, administrative and organisational tasks, information technology	78.2	21.4	0.4	85	0.43
Health care system, insight into costs and contractual procedures	68.3	20	11.7	84	0.66
Total percentage	**74.7**				
Health advocate					
Reflection on aspects influencing general health (bio, psychosocial and existential)	73	24	3	67	0.15
Primary and secondary prevention, patient safety	28.3	58	13.7	61	0.30
Patient's benefits, guidance in health care	65.5	27.5	7	61	0.22
Reflection on critical incidents	69.5	27.5	3	69	0.32
Total percentage	**59.1**				
Scholar					
Scientific thinking and acting	92.9	7.1		88	0.04
Attitude of lifelong learning—a personal development plan	98.8	0.8	0.4	98	–
Transmission of medical knowledge and practice to the scientific community and an evolving society	46.25	53.75		76	0.51
Total percentage	**79.3**				
Professional					
Highest quality of care with an appropriate attitude and behaviour	75.4	24.6		81	0.80
Reflection on professional behaviour and attitude	98.75	1.25		98	–
Total percentage	**87.1**				

The rating categories '+', '+/−', and '−' refer to the ability of portfolio content to assess an item of the VIPA and represent sufficient, moderate and insufficient ability, respectively. *N* is the number of portfolios reviewed. Since two raters reviewed each portfolio, the total number of ratings amounts to $2 \times 120 = 240$. For each item, the category that received the majority of ratings has been indicated in grey. The bold-typed numbers represent the mean percentages of '+' ratings for each role.
Percentage of agreement: percentage of agreement between the three rater pairs as regards the 3-point portfolio classification.
Kappa: Cohen's Kappa for the 3-point classification, weighted mean of Kappa of the three rater pairs (where '−' indicates that Kappa could not be defined).

weighted mean Kappa, using the pair's number of assessed portfolios as weight (Streiner & Norman 2008). Kappa values below 0.4 indicated poor agreement, those between 0.4 and 0.75 signalled sufficient agreement and values beyond 0.75 represented excellent agreement (Fleiss 1981). Finally, we calculated the Percentage of Agreement (PA) within rater pairs per item (Streiner & Norman 2008).

It was not possible to anonymise the portfolios under review. However, we coded the data to ensure confidentiality. Moreover, we obtained ethical approval from the ethics review committees of the Netherlands Association for Medical Education (NVMO) (file number 69, 2011) and of Antwerp University Hospital (file number 2011.023). Students were invited by email to give informed consent.

Results

Analysis of the portfolios' content validity

Table 3 presents the afore specified frequencies (expressed as percentages of total number of ratings per item/role). The portfolio's capacity to assess the different roles ranged from sufficient '(+)' to moderate '(+/−)', with the category 'insufficient (−)' assigned in few instances only, as testified by the low percentages.

Closer inspection of the percentages in the 'sufficient' category revealed that the portfolios the students had composed adequately covered the communicator (90%) and professional (87%) roles. Also the medical expert (82%),

collaborator (82%), scholar (79%) and manager (75%) roles fared well, ranging between 75% and 85%. The health advocate role, however, was less well reflected in portfolio content, although with 59% it was still mirrored in the majority of portfolios.

Percentages per item show that the first four items of the medical expert role are clearly reflected in the portfolio content. An exception to this pattern is the last item 'integration of the different CanMEDS roles' that received a substantially lower rating suggesting that it is more difficult to assess this item using portfolios. Similarly, the items 'primary and secondary prevention, patient safety' (health advocate role) and 'transmission of medical knowledge and practice to the scientific community and an evolving society' (scholar role) were rated as being moderately reflected in portfolio content in over half of the cases.

Analysis of the VIPA's interrater agreement

Table 3 showcases the PA within rater pairs and respective Kappas concerning the 3-point classification of VIPA items. Interrater agreement was found to be 'excellent' for four items ('acquired skills', 'communication of patient cases', 'appropriate contribution in a team of health care providers' and 'highest quality of care with an appropriate attitude and behaviour'); 'sufficient' for six items ('integration of the CanMEDS roles', 'teamwork tasks', 'reflection on self-care and balance between work-personal development', 'priorities, administrative and organisational tasks, information technology', 'health care system, insight into costs and contractual procedures' and 'transmission of medical knowledge and practice'); and 'poor' for eight items ('communication in the doctor–patient[-family] relationship', reflection on own communication', 'reflection on team participation', 'reflection on aspects influencing general health', 'primary and secondary prevention, patient safety', 'patient's benefits, guidance in health care', 'reflection on critical incidents' and 'scientific thinking and acting'). Besides these 18 items, there were seven additional items whose Kappa could not be defined ('medical knowledge', 'medical decision-making', '[multidisciplinary] health care plan', 'intake and history taking', 'communication of scientific research', 'attitude of lifelong learning—PDP' and 'reflection on professional behaviour (PB) and attitude'). These were all items that exhibited both extremely high agreement (PA 98–100%) and a highly dominant class ('+': 96–100%). Obviously, the data for these items present almost perfect agreement, yet too little variance to be able to calculate Kappa.

Discussion

In the present study, we have sought to investigate the content validity of medical students' clerkship portfolios in two European countries. We found evidence confirming that a portfolio can be used as an instrument to evaluate students' achievements in clerkships. This was particularly the case for the roles of communicator, professional, medical expert, collaborator, scholar and manager (in over 75% of the cases, the portfolio provided sufficient information to assess the respective roles). The items pertaining to the health advocate role, however, were more difficult to assess on the basis of portfolio content. When considering the VIPA items individually, we found that the portfolio content sufficiently covered most of them, three items excepted, specifically: 'integration of the CanMEDS roles', 'primary and secondary prevention, patient safety' and 'transmission of medical knowledge and practice to the scientific community'. Nevertheless, these latter items were still rated as being moderately reflected in portfolio content. Hence, from a content validity perspective we may conclude that portfolios can, indeed, be wielded successfully to assess workplace-based learning (Royal College of General Practitioners 1993; Driessen et al. 2007). This outcome finds resonance with previous studies demonstrating that portfolios are helpful in evaluating and steering particular competences, such as the acquisition of required knowledge, communication and critical thinking skills, the ability to work on personal and professional development, recognition of values, building of confidence and the ability to reflect (Mathers et al. 1999; Gordon 2003; O'Sullivan et al. 2004; Rees & Sheard 2004; Driessen et al. 2006; Roberts et al. 2006; McCready 2007; Driessen et al. 2007; Amsellem-Ouazana et al. 2007; Kear & Bear 2007). To our knowledge, only one study to date has conducted similar research by collecting portfolio data from one institution using a self-constructed and validated inventory in order to assess reflection (Driessen et al. 2006). However, the present study should be the first to evaluate the content validity of portfolios with respect to *all* the CanMEDS-based competencies using a validated competency-based inventory.

The health advocate role by all means surfaced as the most difficult one to evaluate. This was evidenced not only by the poor portfolio ratings, but also by raters' divergence of opinion: the portfolio was perceived to contain only moderate information that allowed assessment of the items pertaining to this role, whereas interrater agreement on these very items appeared insufficient. The interrater disagreement on these items related to all three rating categories: sufficient, moderate and insufficient information to enable satisfactory assessment of the item. These findings echo those of previous studies (Chou et al. 2008; Mu et al. 2011; Dobson et al. 2012; Puddester et al. 2015). In a recent publication Puddester et al. (2015), for instance, reported that the 'health advocate emerged as being the most difficult to teach and evaluate', which our results confirm.

Raters tended to disagree on four other items as well, specifically 'reflection on own communication', 'reflection on team participation', 'scientific thinking and acting' and 'communication in doctor–patient(-family) relationship'. Since disagreement stayed confined to the two rating categories 'sufficient information' and 'moderate information', however, the relevance of these results might be negligible. The same holds true for the seven items with undefined Kappas: with almost all portfolios providing sufficient information to assess these items, the sample lacked the variation needed to calculate a reliable level of agreement.

The portfolios involved are competence-based and align with the curricular context and the workplace setting. Consequently, one may draw the inference that achieving content validity is largely a matter of attuning the portfolio blueprint accordingly and relating it to the teaching and

assessment aims. Such assumption ties in with constructive alignment theory (Biggs & Tang 2011) and earlier statements about the quality of assessment instruments, including portfolio (Friedman Ben David et al. 2001; Carraccio & Englander 2004; Michels et al. 2010). Hence, to enhance content validity, education designers would do well to select those portfolio elements that clearly reflect specific competences or roles. It is true that certain elements, such as mini-CEXs, competence development forms or clerkship assessments, can be adjusted easily to suit the required roles or competences. Yet, it has been demonstrated that a careful advance specification of topics can raise the capacity of portfolio elements to assess specific roles and competences.

By extension, the present study calls on curricula and portfolio designers to explore other and new assessment tools or tasks for portfolios. We make this plea because our study revealed that some VIPA items, in particular 'integration of the CanMEDS roles' and 'transmission of medical knowledge and practice', were insufficiently reflected in the content of our clerkship portfolios and, as a corollary, could not be assessed properly. The outcomes also offered compelling evidence that there is a need for new methods to assess all items pertaining to the health advocate role. These could be assignments related to public health or disease prevention. We expect the present update to the CanMEDS Roles Framework to deliver us new insights regarding the different roles and even better alignment with competency-based learning and assessment (Frank et al. 2015). This certainly applies to the health advocate role that according to the CanMEDS Health Advocate Expert Working Group also 'required greater clarity' (Sherbino et al. 2014). Hence, the definition and description of the role will be refined, as a result of which certain competences and concepts will change, whereas new ones, such as 'health equity', will be introduced. Also of interest is the introduction of milestones, which are 'descriptions of the abilities expected of a trainee or physician at a defined stage of professional development' (Frank et al. 2014), and of Entrustable Professional Activities (EPAs), i.e. 'tasks or activities that must be accomplished' (ten Cate et al. 2010; Royal College of Physicians and Surgeons of Canada 2015). These concepts can help to take into account 'the constraints and characteristics of the local environment' when assessing competences and roles.

What may also be inferred when considering the contextual features of the portfolios reviewed (Table 1) is that appropriate assessment procedures, as well as feedback, progress testing and coaching can help bolster content validity. Although the real impact of mentorship and coaching on student outcomes remains largely under-investigated (Tochel et al. 2009), previous research on portfolios has demonstrated that student coaching and progress testing can contribute to a portfolio's success: they encourage students to engage in deeper learning, enhance their reflective ability, better face emotions and plan learning trajectories, all of which are effects that are likely to facilitate the achievement of competences (Challis 1999; Friedman Ben David et al. 2001; Driessen et al. 2007; Epstein 2007; Buckley et al. 2009).

This study is not without limitations. Firstly, it focused on one dimension of validity—content validity—only. As mentioned in the background section, validity encapsulates other dimensions or 'sources/aspects of validity' as well. Hence, to obtain a comprehensive view of the validity of portfolio assessment, one may want to build on the framework proposed by Messick (1995) (Downing 2003) or the approach outlined by Kane (2001, 2006a, b). A second possible limitation is the fact that the Antwerp portfolio was originally premised on the Scottish three-circle model, and, consequently, one may question whether its inclusion in this validity study was appropriate. However, the Antwerp curriculum has transitioned to the CanMEDS Roles Framework, which is reflected in recent portfolio changes. This study therefore provided an opportunity to investigate whether the portfolio is still consistent with the new curriculum. We encourage future studies to replicate the present research using other workplace-based portfolios so as to find out whether our conclusions can be generalised. This could also help to further consolidate the VIPA.

Conclusions

This study supports the growing body of evidence that a portfolio can be a useful tool for assessing the CanMEDS roles and competencies. The health advocate role, however, was less well represented in portfolio content. We established that careful advance determination of the portfolio blueprint and aims, while selecting the appropriate assessment instruments that fit the intended competencies, can greatly enhance content validity.

Notes on contributors

NELE R. MICHELS, MD, PhD, Teacher Skills Lab and Centre for General Practice, Faculty of Medicine, Antwerp, Belgium.

MARIJKE AVONTS, MD, Teacher Skills Lab, Faculty of Medicine, Antwerp, Belgium.

GRIET PEERAER, PhD, Professor, Faculty of Medicine, Antwerp, Belgium.

KRIS ULENAERS, MSc, Dean's office, Faculty of Medicine, Antwerp, Belgium.

LUC F. VAN GAAL, MD, PhD, Full professor, Faculty of Medicine, Antwerp, Belgium.

LEO L. BOSSAERT, MD, PhD, Full professor emeritus, Faculty of Medicine, Antwerp, Belgium.

ERIK W. DRIESSEN, PhD, Full professor, Faculty of Medicine, Maastricht, The Netherlands.

ARNO M. MUIJTJENS, PhD, Statistician-methodologist, Associate professor, Faculty of Medicine, Maastricht, The Netherlands.

BENEDICTE Y. DE WINTER, MD, PhD, Professor, Faculty of Medicine, Antwerp, Belgium.

Acknowledgments

We wish to express our gratitude to all the experts who cooperated in this study. We also wish to thank the students for giving informed consent and hence granting access to their portfolios. In addition, we are grateful to the staff of the medical faculties of Maastricht University and Utrecht University who engaged in the validity study, in particular

Prof. O. ten Cate, Dr M. Van den Acker and Dr T. van Haeften. The authors wish to thank Angelique van den Heuvel for editing the manuscript.

Declaration of interest: The authors report no declarations of interest.

References

Amsellem-Ouazana D, Van Pee D, Godin V. 2007. Utilisation du portfolio au cours du deuxième cycle des études médicales dans un service d'urologie [Use of a portfolio during the second cycle of medical studies in a urology department]. Prog Urol 17:978–982.

Biggs J, Tang C. 2011. Teaching for quality learning at university. Buckingham: Open University Press/McGraw Hill.

Braun V, Clarke V. 2006. Using thematic analysis in psychology. Qual Res Psychol 3:77–101.

Buckley S, Coleman J, Davison I, Khan KS, Zamora J, Malick S, Morley D, Pollard D, Ashcroft T, Popovic C, Sayers J. 2009. The educational effects of portfolios on undergraduate student learning: A Best Evidence Medical Education (BEME) systematic review. BEME Guide No. 11. Med Teach 31:282–298.

CanMEDS. 2000. Extract from the CanMEDS 2000 project societal needs working group report. Med Teach 22:549–554.

Carraccio C, Englander R. 2004. Evaluating competence using a portfolio: A literature review and Web-based application to the ACGME competencies. Teach Learn Med 16:381–387.

Challis M. 1999. AMEE guide n°11 (revised): Portfolio-based learning and assessment in medical education. Med Teach 21:370–386.

Chou S, Cole G, Mclaughlin K, Lockyer J. 2008. CanMEDS evaluation in Canadian postgraduate training programmes: Tools used and programme director satisfaction. Med Educ 42:879–886.

Colbert CY, Ownby AR, Butler PM. 2008. A review of portfolio use in residency programs and considerations before implementation. Teach Learn Med 20:340–345.

Davis MH, Friedman Ben-David M, Harden RM, Howie P, Ker J, Mcghee C, Snadden D. 2001. Portfolio assessment in medical students' final examinations. Med Teach 23:357–366.

Dobson S, Voyer S, Regher G. 2012. Perspective: Agency and activism: Rethinking health advocacy in the medical profession. Acad Med 87:1161–1164.

Downing SM. 2003. Validity: On the meaningful interpretation of assessment data. Med Educ 37:830–837.

Driessen EW, Overeem K, Van Tartwijk J, Van Der Vleuten CP, Muijtjens AM. 2006. Validity of portfolio assessment: Which qualities determine ratings? Med Educ 40:862–866.

Driessen EW, Van Tartwijk J, Govaerts M, Teunissen P, Van Der Vleuten CPM. 2012. The use of programmatic assessment in the clinical workplace: A Maastricht case report. Med Teach 34:226–231.

Driessen EW, Van Tartwijk J, Van Der Vleuten C, Wass V. 2007. Portfolios in medical education: Why do they meet with mixed success? A systematic review. Med Educ 41:1224–1233.

Epstein RM. 2007. Assessment in medical education. N Engl J Med 356:387–396.

Fleiss JL. 1981. Statistical methods for rates and proportions. New York: Wiley.

Frank J, editor. 2005. The CanMEDS 2005 physician competency framework. Better standards. Better physicians. Better care. Ottawa: The Royal College of Physicians and Surgeons of Canada.

Frank J, Snell L, Sherbino J, editors. 2015. The draft CanMEDS 2015 physician competency framework – Series IV. Ottawa: The Royal College of Physicians and Surgeons of Canada.

Frank J, Snell L, Sherbino J,et al. 2014. Draft CanMEDS 2015 milestones guide. Ottawa.

Friedman Ben David M, Davis MH, Harden RM, Howie PW, Ker J, Pippard MJ. 2001. AMEE Medical Education Guide No. 24: Portfolios as a method of student assessment. Med Teach 23:535–551.

Gordon J. 2003. Assessing students' personal and professional development using portfolios and interviews. Med Educ 37:335–340.

Harden RM, Crosby JR, Davis MH, Friedman M. 1999. AMEE guide no. 14: Outcome-based education: Part 5 – From competency to meta-competency: A model for the specification of learning outcomes. Med Teach 21:546–552.

Haynes SN, Richard DCS, Kubany ES. 1995. Content validity in psychological assessment: A functional apporach to concepts and methods. Psychol Assess 7:238–247.

Kane M. 2001. Current concerns in validity theory. J Educat Meas 38:319–342.

Kane M. 2006a. Content-related validity evidence. In: Downing SM, Haladyna TM, editors. Handbook of test development. Mahwah, NJ: Lawrence Erlbaum Associates.

Kane M. 2006b. Validation. In: Brennan R, editor. Educational measurement (4th ed.). Westport, CT: American Council on Education/Praeger.

Kear ME, Bear M. 2007. Using portfolio evaluation for program outcome assessment. J Nurs Educ 46:109–114.

Mathers NJ, Challis MC, Howe AC, Field NJ. 1999. Portfolios in continuing medical education – effective and efficient? Med Educ 33:521–530.

Mccready T. 2007. Portfolios and the assessment of competence in nursing: A literature review. Int J Nurs Stud 44:143–151.

Messick S. 1995. Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. Am Psychoh 50:741–749.

Michels NR, Denekens J, Driessen EW, Van Gaal LF, Bossaert LL, De Winter BY. 2012. A Delphi study to construct a CanMEDS competence based inventory applicable for workplace assessment. BMC Med Educ 12:86.

Michels NR, Driessen EW, Muijtjens AM, Van Gaal LF, Bossaert LL, De Winter BY. 2009. Portfolio assessment during medical internships: How to obtain a reliable and feasible assessment procedure? Educ Health 22:313.

Michels NR, Van Gaal LF, Bossaert LL, De Winter BY. 2010. Portfolio in medisch onderwijs: Ontwikkeling en eerste ervaringen. Tijdschr Geneeskd 66:386–393.

Miller GE. 1990. The assessment of clinical skills/competence/performance. Acad Med 65:63–67.

Mu L, Shroff F, Dharamsi S. 2011. Inspiring health advocacy in family medicine: A qualitative study. Educ Health 24:534.

O'Sullivan P, Reckase M, Mcclain T, Savidge M, Clardy J. 2004. Demonstration of portfolios to assess competency of residents. Adv Health Sci Educ 9:309–323.

Pitts J, Collin C, Thomas P. 2001. Enhancing reliability in portfolio assessment: "shaping" the portfolio. Med Teach 23:351–356.

Puddester D, Macdonald CJ, Clements D, Gaffney J, Wiesenfeld L. 2015. Designing faculty development to support the evaluation of resident competency in the intrinsic CanMEDS roles: practical outcomes of an assessment of program director needs. BMC Med Educ 15:100.

Rees CE. Sheard CE. 2004. The reliability of assessment criteria for undergraduate medical students' communication skills portfolios: the Nottingham experience. Med Educ 38:138–144.

Roberts C, Cromarty I, Crossley J, Jolly B. 2006. The reliability and validity of a matrix to assess the completed reflective personal development plans of general practitioners. Med Educ 40:363–370.

Royal College of General Practitioners. 1993. Portfolio-based Learning in General Practice: Report of a Working Group on Higher Professional Education: December 1993. Occasional paper 63.

Royal College of Physicians and Surgeons of Canada. 2015. Milestones and EPAs: Practical and manageable teaching and assessment in residency training. [Online]. [Accessed October 2015] Available from http://www.royalcollege.ca/portal/page/portal/rc/canmeds/canmeds(2015)/milestones.

Sherbino J, Bonnycastle D, Côte B, Flynn L, Hunter A, Ince-Cushman D, Konkin J, Oandasan I, Regehr G, Richardson D, Zigby J. 2014. The CanMEDS 2015 health advocate expert working group report. Ottawa: The Royal College of Physicians and Surgeons of Canada.

Simpson J, Furnace J, Crosby J, Cumming A, Evans P, Friedman Ben David M, Harden R, Lloyd D, Mckenzie H, Mclachlan J. et al. 2002. The Scottish doctor – learning outcomes for the medical undergraduate in

Scotland: A foundation for competent and reflective practitioners. Med Teach 24:136–143.

Snadden D, Thomas M. 1998. The use of portfolio learning in medical education. Med Teach 20:192–199.

Streiner D, Norman G. 2008. Health measurement scales: A practical guide to their development and use. Oxford: Oxford University Press.

Ten Cate OT, Snell L, Carraccio C. 2010. Medical competence: The interplay between individual ability and the health care environment. Med Teach 32:669–675.

Tochel C, Haig A, Hesketh A, Cadzow A, Beggs K, Colthart I, Peacock H. 2009. The effectiveness of portfolios for post-graduate assessment and education: BEME guide no. 12. Med Teach 31:299–318.

Appendix 1. CanMEDS competency-based inventory.

CanMEDS role	Key competence
Medical Expert the student	1. Has insight in required medical knowledge with regard to a clinical problem, i.e.: a. applies the acquired knowledge b. applies medical decision making 2. Efficiently applies acquired medical skills with regard to a clinical problem 3. Accomplishes a health care plan: a. Performs a relevant and adequate intake and anamnesis b. Performs an efficient physical or other examination c. Generates a differential diagnosis d. Efficiently gathers, analysis, and interprets data (from anamnesis, physical examination, and technical investigations) e. Generates an accurate diagnosis f. Presents efficient treatment plans 4. Generates an accurate, multidisciplinary health care plan with specific attention for patient's self-care and follow-up care 5. Defines symptoms of the most common and critical diseases and recognises alarm symptoms (also for differential diagnosis) 6. Integrates the different CanMEDS roles
Communicator the student	7. Clearly and understandably reports a relevant, complete, systematic and accurate intake and anamnesis 8. Writes reports concerning patients encounters in the medical record and in referral letters to other health care providers 9. Can manage a patient record, and clearly and structurally provides (all) the information to other health care providers. 10. Verbally reports on patients encounters to other doctors and health care providers 11. Communicates scientific research in a clear, complete and structural way 12. Communicates during a patient's encounter according to the rules of good practice 13. Establishes (and maintains) an empathic, trustful and ethical doctor-patient relationship and doctor-family relationship 14. Reflects on own communication skills and his/her progression
Collaborator the student	15. Knows and involves the profile and competences of other health care providers 16. Actively takes part in team work 17. Effectively contributes to the interdisciplinary teamwork concerning patient care, education and research 18. Integrates following aspects with regard to team work: a. Taking and giving responsibility b. Delegating and organising c. Giving and taking suggestions to/of other health care providers d. Supporting the ''chain-of-care'' (increasing effective team work) e. Coping with conflicts between professionals 19. Reflects on teamwork and on respecting the opinions of other team members
Manager the student	20. Reflects on self-care and the balance between work and personal development (work/private time management) 21. Ranks information in order of importance and urgency; responsibly prioritises, and motivates priorities (professional time management) 22. Correctly and punctually deals with administrative and organisational tasks 23. Registers, classifies, and transfers patient related data in an effective (and trustful) way 24. Uses information technology to: a. Optimise patient care and practice organisation – (patient related databases) b. Stimulate ''lifelong learning'' – (medical databases) 25. Can work within the health care system and other care systems (welfare, justice) in Belgium 26. Has insights in costs of medical care and their implication for society, patients and medical doctors 27. Has insights in procedures for solicitations and contractual negotiations
Health advocate the student	28. Reflects on: psychological, social, economical, biological, ethical, cultural, and religious aspects influencing patients' health 29. Attends to the individual patient and the population regarding health-related aspects (primary prevention) 30. Deals with prevention and health promotion for the individual patient and the population (secondary prevention) 31. Has attention for patient safety 32. Efficiently accompanies patients through the health care system and reasons in support of a decision making 33. Prioritises the patient's benefits 34. Involves and facilitates the accessibility of health care during daily practice, especially for vulnerable groups of patients 35. Reflects on critical incidents in doctor's practice

Scholar the student	36. Poses relevant, practical and scientific questions with regard to patient care 37. Performs searches in medical scientific databases/sources in an efficient, purposeful and rapid way 38. Questions the quality of consulted medical scientific databases/sources 39. Adequately applies scientific information in decision making in doctor's practice 40. Development and follow up of a personal learning plan a. Can critically reflect on daily performance in the doctor's practice b. Describes and analyses own personal learning needs c. Applies an adequate learning method d. Self-evaluates or evaluates with peers his learning results and remediates 41. Assists in creating, spreading, and applying new medical knowledge and practice 42. Stimulates training of patients, family, students, trainees, other health care workers, population 43. Adapts his functioning to societal evolutions in health care 44. Is open-minded towards ''lifelong learning''
Professional the student	45. Utilises the highest quality of care for his/her patient in an integral, upright and ethical way 46. Understands the meaning of and applies: a. Professional codes b. Ethical codes and dilemmas (= uses an ethical frame) c. Legal codes 47. Reflects on a. Own behaviour b. Own professional attitude: shows willingness to offer medical care in an optimal, ethical, and patient centred way c. Attitude and behaviour of others and evaluates this for himself/herself d. Legal implications of patient care (patient rights, professional secrecy or professional confidentiality, DNR-codes, end-of-life coaching) e. Professional, ethical and legal codes 48. Has an appropriate professional attitude and behaviour, demonstrating a. Honesty b. Integrity c. Engagement d. Respect e. Understanding, empathy f. Altruism and remediates (him-/herself) when needed 49. Recognises his/her own limits, weaknesses or lacunas and can cope with these