


Stakeholder perspectives on workplace-based performance assessment: towards a better understanding of assessor behaviour

Laury P. J. W. M. de Jonge¹  · Angelique A. Timmerman¹ ·
Marjan J. B. Govaerts² · Jean W. M. Muris¹ ·
Arno M. M. Muijtjens² · Anneke W. M. Kramer³ ·
Cees P. M. van der Vleuten²

Received: 7 July 2016 / Accepted: 24 January 2017 / Published online: 2 February 2017
© The Author(s) 2017. This article is published with open access at Springerlink.com

Abstract Workplace-Based Assessment (WBA) plays a pivotal role in present-day competency-based medical curricula. Validity in WBA mainly depends on how stakeholders (e.g. clinical supervisors and learners) use the assessments—rather than on the intrinsic qualities of instruments and methods. Current research on assessment in clinical contexts seems to imply that variable behaviours during performance assessment of both assessors and learners may well reflect their respective beliefs and perspectives towards WBA. We therefore performed a Q methodological study to explore perspectives underlying stakeholders’ behaviours in WBA in a postgraduate medical training program. Five different perspectives on performance assessment were extracted: Agency, Mutuality, Objectivity, Adaptivity and Accountability. These perspectives reflect both differences and similarities in stakeholder perceptions and preferences regarding the utility of WBA. In comparing and contrasting the various perspectives, we identified two key areas of disagreement, specifically ‘the locus of regulation of learning’ (i.e., self-regulated versus externally regulated learning) and ‘the extent to which assessment should be standardised’ (i.e., tailored versus standardised assessment). Differing perspectives may variously affect stakeholders’ acceptance, use—and, consequently, the effectiveness—of assessment programmes. Continuous interaction between all stakeholders is essential to monitor, adapt and improve assessment practices and to stimulate the development of a shared mental model. Better understanding of underlying stakeholder perspectives could be an important step in bridging the gap between psychometric and socio-constructivist approaches in WBA.

✉ Laury P. J. W. M. de Jonge
l.dejonge@maastrichtuniversity.nl

¹ Department of Family Medicine, FHML, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands

² Department of Educational Research and Development, FHML, Maastricht University, Maastricht, The Netherlands

³ Department of Family Medicine, Leiden University, Leiden, The Netherlands

Keywords Assessment perceptions · Assessor variability · Competency based medical education · Q methodology · Workplace-based assessment

Introduction

Workplace-Based Assessment (WBA) plays a pivotal role in present-day competency-based medical curricula. WBA essentially fulfils two functions: it serves a summative purpose, to enable decisions on the learner's achievement, as well as a formative purpose, in order to drive learning and monitor personal development (van der Vleuten et al. 2010). The authenticity of the clinical environment implies that WBA is complex and typically influenced by uncontrolled variables such as case difficulty, patient mix and numbers. Moreover, validity in WBA mainly depends on how stakeholders (e.g. clinical supervisors, learners, programme directors) use the assessments—rather than on the intrinsic qualities of instruments and methods used to evaluate clinical performance (van der Vleuten and Verhoeven 2013). As a consequence, the utility of WBA is questioned regularly due to issues related to stakeholders' behaviours in the assessment process.

Attempts to enhance the utility of WBA obviously target the quality of assessor judgements. There is a plethora of literature on assessor variability—and implications for WBA validity. Research findings reveal that assessor behaviours are quite persistent despite training and the idiosyncratic nature of assessor judgements may result in large differences between performance ratings (Cook et al. 2009; Govaerts et al. 2007, 2013; Holmboe et al. 2004). More specifically, findings indicate that a broad range of factors may underlie assessor variability, including cognitive factors (e.g. differences in leniency and stringency, stereotyping, categorisation) and the peculiarities of the (social or organisational) assessment context (Cook et al. 2010; Gawronski et al. 2003; Gingerich et al. 2011; Harasym et al. 2008; McManus et al. 2006; Sackett et al. 2002; Yeates et al. 2013). Within the current psychometric discourse in medical education, assessor variability is often seen as unwanted bias or error and assessment practices typically strive to objectify performance and to reach generalisable, reproducible judgements. However, it is increasingly being argued that there can be no such thing as 'objective' evaluation of performance (Newell and Shanks 2014). Taking a constructivist, sociocultural point of view, Govaerts et al. (2007), for instance, propose that assessors should be seen as active and goal-directed information processors, rather than passive measurement instruments (Govaerts et al. 2007).

The complexity of assessor judgements in WBA is clearly reflected in a model by Kogan et al. (2011), presenting multiple factors that explain the variability in judgements of trainee performance (Kogan et al. 2011). According to this model, assessors are not only driven by different frames of reference when observing and rating performance, they also use variable and therefore capricious approaches to translate judgements into numerical ratings. Moreover, assessors interpret trainee behaviour—for example exuding confidence or specific body language- and consequently make subjective inferences and assumptions about trainee performance. The model further suggests that assessment outcomes are also influenced by external factors, such as the clinical context of the observed encounter, the assessor-learner relationship and the (expected) response to- and acceptability of- feedback of both assessor and learner. Finally, there is an acknowledged role of the broader institutional culture in guiding assessor ratings; assessors' beliefs about—and trust in—the assessment system seem to be crucial elements in the utility of assessment systems (Kogan et al. 2011). Performance assessment in workplace settings, then, is to be considered a 'socially situated interpretive act', in which a broad range of social and cognitive factors

interact to produce idiosyncratic individual judgements of performance (Shay 2004). This latter postulate was reinforced by Berendonk et al. (2013), who pointed to the importance of assessors' perspectives on assessment tasks and how these perspectives may influence assessment outcomes. More specifically, their study indicated that assessor behaviours are determined by (various) beliefs about assessment purposes and the utility of assessment for learning. These findings in medical education are consistent with those from research in industrial and organisational psychology indicating that performance ratings may be distorted by beliefs and perspectives that assessors have about the process of performance appraisal (Tziner et al. 1998, 2001, 2005).

Similarly, learners' acceptance of work-based assessments and their use of feedback for competency development is not self-evident. In fact, a large body of research strongly suggests that learners' beliefs about learning and learning outcomes filter and may even distort the message that feedback is intended to convey (Butler and Winne 1995). As such, different beliefs about the goals and meaning of feedback following performance evaluations in real-life workplace settings may impair learners' acceptance and use of feedback (Embo et al. 2010). Teunissen et al. (2013) described the role of self-theories and their associated goal orientations in understanding the motivation underlying learners' feedback-seeking behaviours—and thus the stance learners may adopt in WBA (Teunissen and Bok 2013): For example learners who are oriented towards learning goals regard feedback as useful information that helps to correct errors and achieve mastery; Performance-oriented learners, by contrast, tend to take feedback as a judgement about the self and as an indicator of inadequate ability. Especially when this judgement is perceived as negative, the conception of relevance and usage of feedback can be impaired. Finally, learners' beliefs and attitudes towards fairness of assessments have been demonstrated to have an effect on the acceptance of feedback (Watling and Lingard 2012).

Current research on assessment in clinical contexts thus seems to imply that differing behaviours of both assessors and learners may well reflect their respective beliefs, perspectives and attitudes about WBA. This is well in line with theoretical frameworks that explain strong relationships between a person's beliefs and intentions influencing actual behaviours, such as Ajzen and Madden's Theory of Planned Behaviour (Ajzen 2002; Ajzen and Madden 1986). As a consequence, effective improvement of WBA may first and foremost require better understanding of stakeholders' beliefs and perspectives. In the medical education realm, however, perspectives underlying behaviours in work-based assessment have received scant attention. The purpose of the present study is therefore to explore perspectives underlying stakeholders' behaviours in WBA. Awareness of and knowledge on the content of underlying stakeholder perspectives may help us further enhance the utility and quality of performance assessment in competency-based medical education (CBME). To identify and describe key stakeholders' perspectives regarding performance assessment in workplace settings, we used Q methodology, a well-known method for the systematic investigation of people's viewpoints, beliefs and opinions regarding a certain topic (Watts and Stenner 2012).

Method

Context

This study was conducted at two General Practice (GP) Specialty Training institutes in the Netherlands. These GP programmes have an extensive track record in direct observation

and feedback as cornerstones of competency-based education and assessment. The 3-year postgraduate training programme in the Netherlands consists of 2 years of training in general practice (years 1 and 3) and 1 year (year 2) of rotations in hospitals, mental health institutes and institutions for care of the elderly. Trainees spend 4 days in general practice and return to the training institute for a 1-day release programme every week. Throughout the training programme, a variety of formative and summative assessment methods are used periodically across all levels of Miller's pyramid to evaluate the competency development of trainees. At the workplace, single encounter assessments (e.g. mini CEX, direct observations, case based discussions) are used for daily feedback and as input for an aggregated assessment portfolio based on the CanMEDS framework. This instrument is used as input for comprehensive competence assessment by supervisor, GP trainer and a psychologist teacher.

Methodology

We performed this study using Q methodology, which fits well with our purpose of identifying and clarifying salient similarities and differences between various perspectives on WBA among stakeholders (Brown 2009; Cross 2005). Q methodology combines aspects of qualitative and quantitative research approaches and has successfully been applied in studies in health services (Harvey et al. 2013; Honey et al. 2013; Shabila et al. 2014; Stenner et al. 2003) and medical education (Fokkema et al. 2014; Ha 2014; Meade et al. 2013; Wallenburg et al. 2010). As described in the next paragraphs, Q methodology comprises four stages: (1) definition of the concourse surrounding a certain topic (i.e., WBA) and development of a statement set (Q set) based on the concourse; (2) identification of participants (P set); (3) ranking of statements (Q sort) by participants; and (4) statistical factor analysis, resulting in correlated clusters of Q sorts. These clusters can be interpreted as differing perspectives on the concourse—in our case differing stakeholder perspectives on performance assessment which will be described in the results section of this article (Van Exel 2005).

Concourse definition and development of the Q set

In Q methodology, the flow of communicability surrounding any topic is referred to as a 'concourse', and it is from this concourse that a sample of statements, the Q set, is subsequently drawn to enter a Q sort (Watts and Stenner 2012). In order to ensure coverage of all the relevant ground we developed our Q set through a conceptual review of the recent literature on work-based performance assessment in CBME. Key themes in the literature were identified and discussed iteratively within the research team. In addition, interviews were held with two experts with an extensive international track record in medical education research and two heads of a GP training institute in the Netherlands as a cross-check for the appropriate identification of key themes. Based on the conceptual literature review and interviews, we were able to identify three *key themes* in the concourse: 'psychometric versus social constructivist approaches', 'holistic versus analytic conceptualisations of competence' and 'assessment for learning versus assessment of learning'.

The main researcher (LJ) formulated statements to represent the three themes, producing an initial set of 72 statements. LJ, AT and MG subsequently commented on the ambiguity, clarity and suitability of the statements in an iterative process, resulting in a pilot set of 52 statements. This Q set was then pilot-tested by three research team members (AT, JM, AK) and four potential participants (one GP teacher, two GP supervisors, one GP

trainee), who were asked to complete a Q sort and comment on the completeness of the statement set, overlap and the applicability of statements. This resulted in a final set of 48 statements, which was approved by the research team. Finally, the statements were randomly numbered and printed on laminated cards. The complete list of statements used in the Q sort is depicted in Table 1.

Purposive selection of the P set

In Q methodology, participants must represent a broad range of expertise, roles and responsibilities related to the topic under study and specifically be able to express a defined viewpoint about (in our case) work-based performance assessment in CBME. In the Netherlands, the assessment of professional competence in general practice (GP) specialty training involves various stakeholders: a GP supervisor assesses the trainee through day-to-day observations; GP and psychologist teachers observe and provide feedback on different competences during the weekly 1-day release programme; The trainee is actively involved in the assessment process through regular self and peer assessments of competency development; while the programme director is responsible for high-stakes summative decisions based on the trainee's portfolio. Hence, to cover all of the said stakeholder groups, we selected the P set through purposive stratified sampling from the various stakeholders involved in the assessment process. The P set consisted of 48 participants who were equally distributed between the two general practice specialty training institutes (See Table 2).

Q-sorting procedure

The purpose of the study and instructions for completing the Q-sorting task were described on an information sheet, which was handed out to each participant to obtain informed consent. We then asked participants to read through the statements of the Q set and start the Q-sorting procedure by dividing the statements into three piles: agree, disagree and neutral. A sorting grid with an 11-point distribution (-5 to $+5$) was used as a format to rank-order the laminated statement cards (Fig. 1). From the 'disagree' pile, participants were asked to select the two statements they disagreed with most and to place these in the two spaces at -5 (disagree most in the Q-sorting grid). After that, they returned to the disagree pile and continued sorting according to the Q-sorting grid until no statements were left from this pile. A similar process followed for the agree pile, after which the neutral pile was rank-ordered in the remaining middle area. A selection of participants provided verbal comments on the positioning of their Q-sorts which LJ collected immediately following the Q sort.

Statistical factor analysis

We analysed the data from the Q sorts using the PQMethod 2.35 programme (Schmolck 2014). All Q sorts were subjected to by-person factor analysis to uncover patterns in the rankings of statements, under the assumption that a high correlation between the Q sorts of certain participants indicated similarity of their viewpoints (Watts and Stenner 2012). The prominent common viewpoints were revealed in a three-step procedure. In a first step the scree plot corresponding to the Eigenvalues of the principal component analysis of the set of Q sorts was used to decide about the number of factors (common viewpoints) present in the data (Watts and Stenner 2005a). The corresponding number of components with highest eigenvalues were extracted as factors. Subsequently, the coordinate system of the

Table 1 Complete list of 48 Q sort statements and idealised Q sorts for the five factors representing stakeholders' perspectives on performance assessment in GP specialty training

No.	# Statement	Factor: perspectives				
		1: Agency	2: Mutuality	3: Objectivity	4: Adaptivity	5: Accountability
1.	Assessment of competency development in General Practice may only take place in the workplace setting	-5	-2	-2	-5	1 ^a
2.	Acceptance of negative feedback necessitates a relationship of trust between trainee and supervisor	-2	+2	-3	+1	-2
3.	An assessment instrument should allow monitoring of trainee development	+3	+2	+2	+2	+1
4.	Giving feedback is important	+5	+5	+3	+5	+2
5.	In summative assessment, numerical grades are more appropriate than narrative evaluations	-2	-3	-1	0	-3
6.	For high-quality assessment, my experience as an assessor is more important than my experience as a (trainee or) general practitioner	-1	-1	+1	0	-1
7.	As an assessor, I feel appreciated by the training institute	0	-2 ^a	+1	+3	+4
8.	Learners should be able to compensate for poor grades over time	-2	+2 ^a	-3	-3	-4
9.	Assessment practices ensure high-quality patient care by the trainee	+2 ^a	-3 ^a	0	0	+5 ^a

Table 1 continued

No.	# Statement	Factor: perspectives				
		1: Agency	2: Mutuality	3: Objectivity	4: Adaptivity	5: Accountability
10.	Assessment should be based on the trainees' learning goals and, consequently, be tailored to the individual trainee	0	+3 ^a	-3	-1	-1
11.	Knowing whether an assessment is formative or summative is important	0	-1	+3	+2	-1
12.	It is important for a trainee to ask feedback	+4 ^b	+2	0	0	+1
13.	Summative assessments cannot be conducted by the supervisor	-5	0	-2	-2	-3
14.	Assessment should primarily drive trainees' learning process	+4 ^a	0 ^b	+2 ^a	-2	-1
15.	Competencies cannot be evaluated with (numerical) grades	-4	0	0	-4	0
16.	A constructive cooperation between supervisor and trainee interferes with critical assessment practices	-3	-5	0 ^a	-2	-4
17.	Assessment becomes more accurate due to the longitudinal relationship between supervisor and trainee	+1	+4	-1	+4	0

Table 1 continued

No.	# Statement	Factor: perspectives				
		1: Agency	2: Mutuality	3: Objectivity	4: Adaptivity	5: Accountability
18.	Professional tasks are more easily entrusted to a trainee whose range of ideas and practices are similar to those of the GP supervisor	+2	+1	-1 ^b	-4 ^b	0
19.	Assessment interferes with the relationship between supervisor and trainee	-3	-4	-5	-3	-5
20.	A capable trainee is easy to recognise	+1	+1	-2 ^a	+1	+1
21.	A trainee who performed well before may be expected to perform well again	+2	0	0	+1	+2
22.	I am a proficient assessor	+3	+1	+2	+3	+4
23.	Clear and precise assessment criteria are needed to assess a trainee accurately	+2 ^b	-1	+4 ^b	-1	0
24.	Numerical grades are not suitable for formative assessments	-2	-2	+3 ^a	-2	-1
25.	When conducting an assessment, progressive development is more important than actual performance	-1	+3 ^a	0	-3	-2
26.	An experienced supervisor is capable of conducting more accurate assessments	0	+2	+1	+3	+3
27.	As an assessor I feel involved with the training institute	0 ^a	-5 ^a	+1	+2	+2

Table 1 continued

No.	# Statement	Factor: perspectives				
		1: Agency	2: Mutuality	3: Objectivity	4: Adaptivity	5: Accountability
28.	Assessment implies an additional workload	-1	-1	+2	0	+2
29.	If the purpose of the assessment is summative, my evaluations are stricter	-1	-2	-3	-1	0 ^b
30.	Numerical grades allow me to assess accurately	-1	-3	-2	-1	-3
31.	Assessment within GP specialty training contributes to the future quality of general practitioners ^c	+5	+3	+5	+3	+3
32.	Assessment practices stimulate the competency development of trainees	+3	0	+4	0	+4
33.	It is important that a trainee and GP supervisor have shared perspectives on the GP profession	-4	-1	-4	-5	-1
34.	A trainee's perspectives on the profession of general practice affect his/her assessment	+1	-4	-1	1	-3
35.	Assessors should judge in an identical fashion	0	+4	+2	0	-1
36.	Competencies are not to be assessed independently of one another	-3	0	-2	-2	-5 ^b
37.	Summative assessments are more important than formative assessments	-4	-2	-4	-1	-2

Table 1 continued

No.	# Statement	Factor: perspectives				
		1: Agency	2: Mutuality	3: Objectivity	4: Adaptivity	5: Accountability
38.	Professional tasks can be entrusted earlier to a trainee who self-directs his or her learning process	+4	0	0	+2	-2
39.	For the progressive development of competencies a trainee's learning goals are more important than formal assessment criteria	-1	+3	0	0	+3
40.	When assessing a trainee, it is crucial that a trainee can perform professional tasks independently	0	-4	-1	-4	0
41.	When assessing a trainee, clear and precise assessment criteria are more important than the personal opinion of the supervisor	-3	-3	-1	-1	-4
42.	Rigorous assessment requires that both trainee and supervisor can receive feedback	+3	+5	+3	1 ^a	+5
43.	Trainees are more likely to learn from narrative assessments than from numerical grades	+2	+4	+4	+4	+1
44.	In summative decisions previous formative assessments should not be taken into consideration	-2	0	+1 ^b	-2	-2
45.	Previous experiences with this trainee influence my assessment	+1	+1	-4 ^a	+5 ^b	+2

Table 1 continued

No.	# Statement	Factor: perspectives				
		1: Agency	2: Mutuality	3: Objectivity	4: Adaptivity	5: Accountability
46.	My style of giving feedback is influenced by the way I expect it to be received	0	-1	-5 ^a	+4 ^a	0
47.	In the assessment process I include assessments of other assessors ^c	+1	+1	+1	+1	+3
48.	It is important to document assessments regularly	+1	+1	+5 ^b	+2	+1

^a Distinguishing statement ($P < .01$)

^b Distinguishing statement ($P < .05$)

^c Consensus statements (those that do not distinguish between ANY pair of factors, non-significant at $P > .01$)

Table 2 P set representing stakeholders involved in performance assessment in CBME in two General Practice Specialty Training Institutes in the Netherlands

Training institute	Number of participants				
	GP supervisor (workplace)	GP teacher (STI ^a)	Psychologist teacher (STI ^a)	Programme director	GP trainee (1st year/3rd year)
Maastricht	10	3	2	1	4/4
Nijmegen	10	2	3	1	4/4

^a General Practice Specialty Training Institute

factor space was rotated by varimax in order to optimize the loadings of the Q sorts for selection of subsets corresponding with a single factor. Only factors representing at least two Q sorts that exhibited a statistically significant correlation ($P < .05$) and having Eigenvalues greater than 1 were extracted (Shinebourne 2009). Here, the optimal number of factors was five. LJ, AT, MG and AM then examined the interpretability of the factor structures, which outcome confirmed that a five-factor solution indeed provided the most comprehensible fit.

As a next step, we created an idealised Q sort for each of the five factors (Table 1). These Q sorts indicated how a participant with that same perspective would have sorted the statements in the Q sorting grid (Fig. 1). Distinguishing statements (i.e., those statements that exceeded the difference score between any two factors at significance levels $P < .05$ or $< .01$) and consensus statements (i.e., those statements that did not distinguish between any pair of factors) were also identified (Table 1). For a holistic interpretation of the perspectives it is important to consider not only the extreme poles of the sorting but also and especially the relative positioning of the statements across the different factors. We therefore created arrays of differences representing the biggest differences in standard

next paragraphs we will describe each factor (i.e., perspective), the corresponding number of the defining statements (e.g. #15) and their position in the idealised Q sort for that factor (e.g. +4). Each perspective will be clarified by means of an illustrative comment on the Q sort from one of the participants. For the complete list of numbered statements and their positions in the idealised Q sort, see Table 1. The table shows the pattern of level of agreement (on a scale -5 to +5) over all statements for each Factor, thereby defining the characteristics of that specific perspective. It should be emphasized that this table does not show loadings. In the Q factor analysis loadings are defined between participants and factors: participants with a score pattern over statements similar to the score pattern shown in one of the columns of the table show a high loading on that particular factor. The relation between the factor scores in a row is important: it expresses the between-factor differences in agreement for a statement. For each factor a selection of these score data for the subset of statements with a salient/significant score for that particular factor was used to reveal the characteristics of a perspective on WBA. The results of these analyses are presented below.

Perspective 1: agency

This perspective holds that assessment should primarily guide the trainees' learning process (#14: 4) and that feedback is central to learning. Active, self-directed learning on the part of the trainee plays a key role in the assessment process. It is especially important that the trainee actively seek feedback and that the assessor-supervisor provide this feedback (#4:+5; #12:+4). Summative assessments are not considered more important than formative assessments (#37:-4), connoting a commitment to assessment *for* learning and the learning process itself. It is essential that assessment instruments allow monitoring the development of trainees (#3:+3). A self-directed learning style implies that a trainee will ask for help when necessary, making supervisors feel comfortable to entrust professional tasks (#38:4). Both formative and summative performance assessments by the GP supervisor (#13:-5) support continuous assessment and guidance for learning, provided that clear and precise assessment criteria are available (#23:+2). Factor 1 explained 18% of variance and was defined by nine participants, five of which trainees, three GP supervisors and one GP trainer.

Supervisor 9: Most important is that trainees are able to learn, therefore, a test should primarily drive learning. The training institute should provide guidance, however, in the end it is the trainee who decides what is to be learnt. And that is fine with me: a capable trainee is able to do so.

Perspective 2: mutuality

In this view, assessment should be embedded in the learning process and a joint responsibility of trainee and supervisor. To allow rigorous assessment, both trainee and supervisor must have the opportunity to receive feedback and trainees must be able to trust their supervisor (#42:+5; #2:+2). A constructive collaboration between supervisor and trainee is not perceived to interfere with critical assessment (#16:-5). In fact, when built on trust and used for learning, critical evaluation of performance can strengthen the trainee-supervisor relationship; assessment is more about process than it is about outcome (#19:-4). It is held, moreover, that the assessment becomes more accurate due to the longitudinal, mutual relationship between resident and supervisor (#17:+4). Continuity is also important

in the assessment process itself: assessment should focus on progressive development (#25:+3) and learners should be able to compensate for poor grades over time (#8:+2). To guide trainees' competency development towards professional standards, feedback is important (#4:+5). Trainees' learning goals are considered more important than formal assessment criteria; Therefore, feedback should be tailored to trainees' needs and goals (#39:+3; #10:+3). It is believed that trainees are more likely to learn from narrative assessments than from numerical grades (#43:+4), even in summative assessments (#5:-3). Prioritising the GP supervisor-trainee relationship, assessors sharing this perspective felt the least involved with—and appreciated by—the GP training institute compared to the other perspectives (#27:-5; 7:-2). Factor 2 explained 9% of the study variance and was defined by five participants: three trainees and two GP supervisors.

Supervisor 2: As a supervisor is assigned to one trainee for one year, you also have the time to invest in your relationship. By doing so, you are able to gain a clear impression of the trainee's performance, not only of his medical skills, but also of his personal acting. This information is valuable, because it is also useful in general practice: how do you perceive a patient in relation to his context. When the relationship is good, you should be confident that an assessment is fair and that it is meant to drive learning. This only strengthens your relationship.

Perspective 3: objectivity

Holders of this perspective attributed two different, yet equally important roles to assessment in GP specialty training (#37:-4): a formative role [*for* learning (#14:2; #32:+4)] and a summative one [*of* learning, to ensure future high-quality care (#31:+5)]. For the purpose of accountability, regular documentation of assessments by using an audit trail or portfolio was considered highly important (#48:5), as was the role of assessor, which required experience (#6:+1) and included specific tasks involving an additional workload (#28:+2). Assessment should be based on clear and precise criteria (#23:4), be objective and not biased or influenced by previous experiences (#45:-4), expectations (#20:-2), assessment purposes (#29:-3) or interpersonal relationships (#17:-1). Moreover, it should not be tailored to individual learning goals (#10:-3). Feedback must be honest and complete, without exceptions (#46:-5), and may even be unsolicited to advance competency development towards (external) standards (#12:0). The GP supervisor-trainee relationship must not be affected by the assessment (#19:-5), nor is it considered important that the resident and GP supervisor share hold similar perspectives on the GP profession (#33:-4). Numerical grades are not suitable for formative assessment as trainees are likely to learn more from narrative feedback than from numerical grades (#23:+3; #43:+4). Factor 3 explained 12% of the study variance and was defined by five participants: three GP supervisors, one GP teacher and one trainee.

Supervisor 7: 'It is not necessary that one holds similar perspectives on the profession; friction can be productive. Criteria are important; it must be clear what the expectations of the training institute are.'

Perspective 4: adaptivity

This view on performance assessment envisages a more flexible role for the assessor with regard to ownership of the assessment process: The relationship between supervisor and trainee should not necessarily be close (as opposed to the mutuality perspective of factor 2)

and little weight is attached to the sharing of ideas about the GP profession, even when entrusting a trainee with specific professional tasks (#33:−5; #18:−4). Although feedback is allocated a prominent place (#4:+5), for rigorous assessment it is not necessary that both trainee and supervisor can receive feedback, suggesting a focus on one-way feedback delivery rather than a feedback dialogue (#42:1). Yet, assessors adjust their styles to the way they expect the feedback will be received (#46:+4), which stands in stark contrast with perspective 3 (objectivity), as does the acknowledgement that previous experiences influence assessment (#45:+5). While trainees are likely to learn more from narrative assessments than from numerical grades (#43:+4), it is certainly possible to evaluate competencies with grades (#15:−4), as long as this is supported by—and explained in—the training programme.

Additionally, it is believed that assessment becomes more accurate as the supervisor-trainee relationship develops (#17:+4) and as the supervisor accumulates experience (#26:+3). The training institute also fulfils an important role in providing more standardised guidance of the assessment process: assessment of competency development does not have to take place in the clinical setting only (#1:−5) and should not primarily drive a trainees' learning process (#14:−2). This latter view reflects a proclivity towards external regulation which contrasts starkly with perspective 1 (agency). The relatively high scores on involvement with—and appreciation by—the training institute (#27:+2; #7:+3) also illustrate the importance of the role of the GP specialty training institute. Factor 4 explained 9% of the study variance and was defined by three participants: two GP supervisors and one trainee.

Third-year trainee 5: 'You don't need to have similar ideas about the profession. (...) You can be a very different type of general practitioner'.

Supervisor 5: Giving feedback is important. To give feedback to the trainee, I have to rely on my opinions and experience; however, in the end it is the training institute that determines what to do and how you should operate. By doing so, you can keep a 'clean' relationship with your trainee. After all, you depend on each other for one year.

Perspective 5: accountability

From this standpoint, assessment practices do not only stimulate the competency development of trainees (#32:+4), but also serve to ensure high-quality patient care (#9:+5). In comparison with the other perspectives, this view holds that competency development should only be assessed in the clinical setting (#1:+1) and assessors do not experience any difficulties with assessing competencies independently of one another (#36:−5). Supervisors consider themselves proficient assessors (#22:+4) and feel involved with—(#27:+2) and much appreciated by (#7:+4) the GP specialty training institute. They also require experience to improve the accuracy of assessments (#26:+3) and there is room for idiosyncratic manoeuvre: they do not have to judge in an identical fashion (#35:−1) and personal opinions can be more important than (clear and precise) assessment criteria (#41:−4). The assessor is accountable for assessment in a mutual constructive relationship, where both trainee and supervisor can receive feedback (#42:+5). Assessment does not seem to jeopardise the supervisor-trainee relationship (#16:−4), not even in critical assessment practices (#19:−5). Moreover, assessors are demanding in the sense that they expect trainees to show a consistent level of performance over time, compensation over time is certainly not desirable (#8: −4). Finally, low priority is given to monitoring

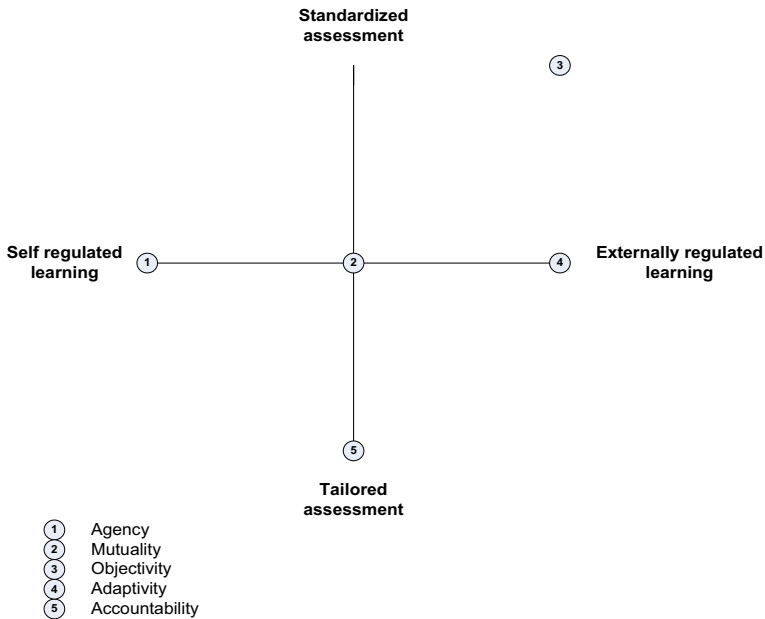


Fig. 2 Conceptual space diagram depicting the positioning of the different perspectives on workplace-based assessment relative to the desired level of self-regulation/externally regulated learning and of standardisation/tailoring of assessment

(#3:+1), indicating that trainees should perform well on each occasion. Factor 5 explained 10% of the study variance and was defined by three participants: two trainees and one GP supervisor.

First-year trainee 1: ‘The two of you have to go for it, otherwise it will not work. A proficient assessor is demanding, and also self-demanding. (...) Therefore, as a trainee, you also need to aim high’.

Discussion

In this study, we used Q methodology to identify and describe stakeholders’ perspectives on WBA in a postgraduate medical specialist training setting. We were able to extract five different perspectives on performance assessment: Agency, Mutuality, Objectivity, Adaptivity and Accountability. These perspectives reflect both differences and similarities in stakeholder perceptions and preferences regarding the use and utility of WBA. In comparing and contrasting the various perspectives, we identified two key areas of disagreement, specifically ‘the locus of regulation of learning’ (i.e., self-regulated versus externally regulated learning) and ‘the extent to which assessment should be standardised’ (i.e., tailored versus standardised assessment). Q methodology often uses a conceptual space diagram as a graphical representation of the different preferences and relationships between the various factors (perspectives) and as a means of identification of the pertinent disagreement areas (Milcu et al. 2014; Stenner et al. 2000; Watts and Stenner 2005a, b). By positioning the various perspectives and viewpoints relative to the axes representing areas

of disagreement, possibilities for comparison and contrast across perspectives can be maximized (see Fig. 2). Whereas the horizontal axis reflects the desired level of self-regulated learning/externally regulated learning, the vertical axis refers to preferred levels of standardisation/tailoring of assessment. The various perspectives are positioned relative to these axes and to one another and are indicated by the numbers 1–5. In the following section, we will discuss the potential implications not only of differing, but also of shared perspectives among stakeholders for the utility of WBA in medical education practice.

The positions of the various perspectives along the horizontal axis in Fig. 2 clearly demonstrate that the importance attributed to self-regulation in learning and assessment differs widely amongst stakeholders. On the left-hand side of this axis, which point coincides with perspective 1 (Agency), stakeholders hold the view that trainees should actively self-direct their learning, take responsibility and show initiative in the assessment process. Halfway the axis, where perspective 2 (Mutuality) is located, the responsibility to identify learning needs through performance evaluations and feedback rests with both supervisor and trainee. Similarly, perspective 5 (Accountability) holds that supervisors must actively engage and take responsibility in the assessment process, and be driven by personal conceptions and beliefs about the role of assessment in ensuring high-quality patient care. At the other end of the spectrum we find both perspectives 3 (Objectivity) and 4 (Adaptivity) that perceive external criteria and requirements as main drivers of learning and WBA.

The differences in desired levels of self- and external regulation of learning in WBA, presented on the horizontal axis of our diagram, resonate with developments and research findings in medical education. Medical education institutions and accrediting regulatory bodies have acknowledged the importance of self-regulated learning for improving academic and clinical performance (Di Genova et al. 2015; Sandars and Cleary 2011). In the recently revised CanMEDS framework, there is a central role for self-regulated learning with regard to assessment: ‘a trainee has to develop, implement, monitor, and revise a personal learning plan to enhance professional practice’ (Frank et al. 2015). At the same time, the outcomes-based approach towards assessment in CBME almost inevitably implies the use of predefined and pre-specified competency frameworks. Detailed assessment criteria, performance standards and milestones may support learners and assessors in making decisions about achievement and inform future learning processes (ten Cate and Scheele 2007). In other words, by serving as a frame of reference, external assessment criteria may provide guidance and inform assessment decisions as well as the direction of future learning. Obviously, stakeholders’ perspectives on assessment, which translate into specific behaviours, will affect the utility of assessment systems.

Research findings strongly suggest that effective WBA implies a joint responsibility between learner and assessor (Holmboe et al. 2010; Norcini and Burch 2007). A prerequisite for effective self-regulated learning is that the supervisor entrusts at least part of the monitoring of learning goals to the learner and creates a learning environment that encourages mutual delivery and receipt of feedback (Pelgrim et al. 2012). In hierarchical learning environments, however, two-way feedback interaction is not self-evident; feedback often remains a supervisor-driven process, as reflected in perspective 5 (Accountability) (Archer 2010). When supervisors use personal conceptions of patient care or external assessment criteria, rather than personalised learning goals as the starting points for assessment activities, self-directed learning and engagement in self-monitoring of competency development can be impaired. On the other hand, excessive and exclusive reliance on learners’ initiatives and responsibility through self-assessments may also have a negative bearing on the utility of assessment processes. In fact, research findings

consistently show that it is difficult to recognise one's own incompetence, making self-assessment not the preferred mechanism to identify areas of personal weakness (Hodges et al. 2001; Regehr and Eva 2006). In this sense, external assessment frameworks can, indeed, serve an important purpose: to help identify learning needs and shortcomings learners have not been able to discover for themselves (Tochel et al. 2009). Yet, there is a caveat in that WBA may become a tick-box exercise if assessment is perceived to be completely driven by externally dictated criteria. Especially learners and assessors who strongly favour individualised assessment approaches will likely lack commitment and a sense of ownership in WBA.

Developers of assessment programmes must take into account the different user perspectives on WBA when combining summative and formative elements to stimulate assessment-for-learning. The effectiveness of assessments may depend on stakeholders' beliefs and their associated perspectives on the assessment process. A lack of clarity regarding the purpose of the assessment, which is to promote self-regulated learning, may very well preclude successful implementation of WBA. In a study on learner-centeredness of a newly developed assessment programme, Bok et al. (2013), for instance, found inconsistencies between the learning function of assessment and its contribution to high-stakes decisions. Assessments that were designed as formative learning experiences were increasingly perceived as summative by learners, thereby undermining self-directed learning activities (Bok et al. 2013). In conclusion, mismatched, but also magnified shared perspectives on self- and externally regulated learning among stakeholders may result in serious impairment of the utility of work-based assessment practices.

The second area where stakeholders' perceptions diverged concerns the preferred levels of standardisation of both assessment instruments and the assessment process itself. The perspectives presented on the vertical axis of the conceptual space diagram reflect these preferences. At the lower end of the vertical axis, perspective 5 (Accountability) holds that assessment must be based in the clinical context (also, contextualised) and tailored to the learning task at hand. Moreover, contextualised task requirements and related assessment criteria are regarded as more important than standardised and de-contextualised assessment criteria or performance standards. Expert judgement of clinical performance is considered crucial in ensuring that assessment is aligned with the requirements of high-quality patient care. Positioned in the middle range of the vertical axis, perspective 4 (Adaptivity) acknowledges the role of the social and organisational context of assessment. More specifically, feedback must be tailored to the characteristics of the clinical context and to the expectations of those involved in patient care. This emphasis on context specificity, however, is counterbalanced by programme developers and directors who are expected to guard the standardisation of the assessment process during medical training. Perspective 1 (Agency), too, recognises that trustworthy and fair high-stakes decision-making requires clear and predefined assessment criteria. At the same time, it holds that both formative and summative assessments should be tailored to the individual learning process. In between perspectives 1 and 4, we find perspective 2 (Mutuality) which, prioritising progressive development over time, prefers that assessment criteria be tailored to the learner's needs and goals. To ensure that the assessment is robust, from perspective 2 care must be taken that assessors arrive at judgements in an identical, reproducible fashion. This focus on the accuracy and validity of assessments—psychometric criteria of assessment quality—is even stronger at the end of the range: Perspective 3 (Objectivity) clearly strives for objectification and standardisation of a performance assessment. Assessors should judge in an identical fashion, and there is a need for predefined clear and precise assessment criteria.

From this perspective, adaptation of assessment to personal learning goals of the learner is considered as the least important.

Tensions arising from these opposing views on the level of standardisation of assessment in WBA resonate well with existing assessment literature. To enhance the accuracy of performance ratings (true scores capturing real, 'objectified' performance) from a quantitative psychometric point of view, the use of uniform test conditions and calibrated assessors is promoted. However, also within the field of psychometrics there is an ongoing debate about appropriate use of assessment instruments and interpretation of assessment results, specifically related to acknowledging the importance of individual differences as well as context (Schoenherr and Hamstra 2016). This is in line with discussions about utility of assessment protocols that can be tailored to the learning task at hand versus those that are pre-structured, detailed and standardised (Eva and Hodges 2012; Schuwirth and Swanson 2013).

In theory, assessors can be allowed more or less latitude in their judgments regardless of whether assessment tasks are selected by trainees as self-regulatory learners or chosen externally. As a matter of practice, however, having standardized assessments by standardized judges implies a limited number of things that can be assessed. Conversely, allowing trainees to choose whatever they want to be assessed would largely preclude standardized assessment because of constraints in time and effort in developing assessments. Non-standardized assessments that are tailored to the individual learner, however, may be perceived as biased, invalid, even unfair and less reliable than competence assessments in standardised assessment settings (Hodges 2013). The different stakeholder preferences with regard to the level of standardisation are also reflected in the way stakeholders (assessors and learners) perceive the utility of assessment instruments. Murphy (2008), for instance, found that stakeholders had diverging perceptions of the usefulness and qualities—and thus acceptability—of the various assessment formats used in GP specialist training (Murphy et al. 2008). It might be hypothesised that assessors whose primary concern is to eliminate subjectivity in the assessment process and to reach objective standardised judgements tend to prefer extensive and detailed rating scales; Assessors favouring a contextualised and holistic judgement, by contrast, are more likely to appreciate assessment frameworks that allow for tailored, individualised judgements (Kogan et al. 2015). When multiple assessors collectively take assessment decisions, then, these may be based on different perspectives. Moreover, differing personal views on the validity of assessment data can impair the utility of such team decisions. It is therefore crucial that any divergent preferences be spelt out in the decision-making process. If preferences regarding the level of standardisation of assessment criteria are appropriately aligned, assessors can work towards a shared mental model of functions and goals of assessment, and implications for assessment design (e.g. instruments, criteria). In conclusion, awareness of both differing and shared perspectives on assessment practices can increase mutual stakeholder understanding and therefore the utility of the assessment process. Mismatched preferences as to the needed level of standardisation of assessment can be a source of incomprehension and potentially obstruct effective assessment practices.

Strengths and limitations

This study based on Q methodology has several strengths. First, we drew the Q set from a wide, representative range of current concourses in the domain of WBA. Second, our participant group consisted of different key stakeholders, representing the full assessment

process in a general practice specialty postgraduate training setting. Third, our holistic approach to the data, taking into consideration the relative positions of all statements, resulted in a comprehensive and nuanced set of perspectives on WBA (Watts and Stenner 2005a, 2012). Finally, we found a statistically significant correlation between all five perspectives and different stakeholders in WBA. These findings indicate that the variability in perspectives results not only from the role of the stakeholder but also from particular preferences among individual stakeholders. The importance of understanding stakeholders' perspectives on WBA is underpinned by psychological theories linking beliefs and intentions to behaviours (Ajzen and Madden 1986; Cilliers et al. 2015). The various perspectives on workplace-based performance assessment we identified amongst stakeholders may equally translate into different assessor and learner behaviours—fundamental to WBA validity and therefore utility.

There are some limitations to our study. First, the perspectives we described are not to be seen as archetypes for the classification of stakeholders. Individual stakeholders are likely to recognise aspects of several perspectives as their own, and perhaps identify with one of the presented perspectives more than with others. This is substantiated by comments of participants during the member checking procedure [e.g. “What I am trying to say here is, that although I recognise myself most in the perspective of Mutuality, this does not mean, that elements from the other four perspectives are strange to me, in the contrary” (psychologist teacher), or “I feel most comfortable with perspective 5 Accountability. However I do recognise elements of the other four perspectives, but less outspoken” (GP trainee)]. In addition, the aim of Q methodology is to sample the range and diversity of views expressed and not to make claims about the percentage of stakeholders expressing them. (Cross 2005) As a corollary, the generalisability of perspectives to specific sub-groups (e.g. supervisor, trainee) is limited (Watts and Stenner 2005a).

Similarly, the transferability of the present findings to other medical specialties and work-based learning settings may be restricted. Our study was confined to a medical specialty postgraduate training setting characterised by long-term one-to-one contacts between supervisor and learner. In contrast, hospital-based supervisors typically have short-term contacts with multiple trainees. Although in these circumstances becoming aware of one's perspectives on WBA is equally important, the lack of prolonged contact in the learning process may hamper the development of shared perspectives, impairing the utility of WBA. Second, we drew the statement set from a wide, representative range of current concourses in the domain of WBA. However, there are substantive differences between the various forms that WBA takes. WBA is not restricted to a single tool neither is it used to assess a single skill only. A practical consequence of this could be that during the sorting procedure participants occasionally had a specific assessment tool or skill in mind, instead of the larger and more general concourse on WBA. Therefore our statement set consisted of 48 well balanced statements on WBA, representing a variety of tools and purposes.

Implications for practice and research

WBA plays a pivotal role in competency-based medical curricula. The various perspectives and resulting behavioural differences in WBA may well explain why the implementation of competency-based assessment has proved so arduous (Hawkins et al. 2015; Holmboe 2014). Differing perspectives may variously affect stakeholders' acceptance—and, consequently, the effectiveness—of assessment programmes.

Stakeholders in assessment practices should discuss the potential implications of both differing and shared perspectives on the utility of WBA, to avoid illegitimate inferences and interpretations of assessment outcomes. Excessive and exclusive reliance on only one of the stakeholders preferences may have a negative bearing on the utility of assessment processes. Therefore, establishing latitude regarding what is assessed and how it is assessed should be a joint responsibility.

Holmboe (2011) identified training of faculty in both clinical and assessment competence as ‘the missing link in CBME’ (Holmboe et al. 2011). Recent qualitative research by Kogan et al. (2015), moreover, revealed that assessors perceived training to positively influence specific assessment skills and to provide them with an enriched, more granular assessment vocabulary (Kogan et al. 2015). At the same time, these authors identified four factors that inhibited or precluded the application of training merits: some assessors preferred holistic assessments to checklists; they felt unable to define competence despite training; they experienced difficulty in changing their approach to assessment; or they expressed concern that they would be the firsts (that is, a minority) to adopt the new approach in their institution. In a recent review on user perceptions of WBA, Massie et al. (2015) identified three principal shortcomings of current WBA implementation: lack of clarity as to the purpose of WBAs, inadequate training in the provision of quality feedback in WBA and time constraints (Massie and Ali 2015). Although these findings support the need for adequate training of both assessors and learners increased awareness of various underlying perspectives not only by trainers, but also by assessment programme developers, will enhance the sophistication and utility of the assessment process.

Assessor judgements that deviate from the majority interpretation may represent important variants of how the assessment task is perceived (Gingerich et al. 2014). Variation in performance interpretation should be taken into account to arrive at a more comprehensive and coherent picture of a learner’s competence (Gingerich et al. 2011). Stakeholders in the assessment process should therefore be encouraged not only to document their performance interpretations, but also to articulate underlying values and assumptions in order to enhance WBA validity (Govaerts and van der Vleuten 2013). In summary, continuous interaction between all stakeholders is essential to monitor, adapt and improve assessment practices and to stimulate the development of a shared mental model (van der Vleuten et al. 2014). Future work should focus on the relation between specific perspectives and cognitive processing of individual stakeholders in WBA and actual behavioural approaches in assessment. This study has been exploring perspectives on WBA and contained small groups of stakeholders. As we stated, by definition in Q methodology the generalisability of perspectives to specific subgroups (e.g. supervisor, trainee) is limited (Watts and Stenner 2005a). Therefore, future work is needed to elucidate differences between the various stakeholder groups. Continued research in this arena, especially in the form of field or action research, will more clearly delineate the practical consequences of differing stakeholder perspectives on the utility of WBA.

Conclusion

This study may contribute to our knowledge in the emerging field of assessor and learner cognition. It may enhance our understanding of the factors inhibiting and facilitating stakeholders’ acceptance of assessment systems and their trust in them, as well as of the effectiveness of feedback processes in performance assessment. This study indicates that

stakeholders may very well hold different perspectives on goals and functions of WBA, which, in turn, may induce different perceptions of the role and responsibilities of the assessor and learner, assessment purposes, assessment process, and finally, the intended assessment outcome. Awareness and knowledge of stakeholder perspectives may deepen our understanding of stakeholders' behaviours and interactions in assessment systems. Potential tensions amongst stakeholders, ensuing from different perspectives on and beliefs about WBA, mirror areas of concord and discord between prominent research perspectives (Gingerich et al. 2014). Our findings emphasise the importance of researchers and practitioners integrating aspects of different perspectives into a shared view. Awareness and understanding of underlying stakeholder perspectives could be an important step in bridging the gap between psychometric and socio-constructivist approaches in WBA.

Acknowledgements The authors would like to thank Angelique van den Heuvel for critically reading and correcting the English manuscript.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix: Descending array of differences

Factor 1 Descending array of differences^a

Factor 2 $1 > 2$

- 9. Assessment practices assure high quality patient care by the trainee. 2.119
- 23. Clear and precise assessment criteria are needed to assess a trainee accurately. 1.748
- 38. Professional tasks can be entrusted earlier to a trainee who self-directs his or her learning process. 1.602
- 14. Assessment should primarily drive trainees' learning process. 1.571
- 27. As an assessor I feel involved with the training institute. 1.436

$2 > 1$

- 2. Acceptance of negative feedback necessitates a relationship of trust between trainee and supervisor. 1.854
- 35. Assessors should judge in an identical fashion. 1.680
- 8. Learners should be able to compensate for poor grades over time. 1.603
- 25. When conducting an assessment, progressive development is more important than actual performance. 1.500
- 39. For the progressive development of competencies a trainees' learning goals are more important than formal assessment criteria. 1.446

Factor 3 $1 > 3$

- 46. My style of giving feedback is influenced by the way I expect it to be received. 2.066
- 45. Previous experiences with this trainee influence my assessment. 1.971
- 12. It is important for a trainee to ask feedback. 1.734
- 38. Professional tasks can be entrusted earlier to a trainee who self-directs his or her learning process. 1.465
- 10. Assessment should be based on the trainees' learning goals and, consequently, be tailored to the individual trainee. 1.153

$3 > 1$

- 15. Competencies cannot be evaluated with (numerical) grades 1.728
 - 24. Numerical grades are not suitable for formative assessments. 1.611
 - 28. Assessment implies an additional workload. 1.349
 - 35. Assessors should judge in an identical fashion. 1.278
 - 48. It is important to document assessments regularly. 1.258
-

 Factor 1 Descending array of differences^a

Factor 4 $1 > 4$

14. Assessment should primarily drive trainees' learning process. 2.495
18. Professional tasks are more easily entrusted to a trainee whose range of ideas and practices are similar to those of the GP supervisor. 1.941
40. When assessing a trainee, it is crucial that a trainee can perform professional tasks independently. 1.631
23. Clear and precise assessment criteria are needed to assess a trainee accurately. 1.345
12. It is important for a trainee to ask feedback. 1.299

 $4 > 1$

2. Acceptance of negative feedback necessitates a relationship of trust between trainee and supervisor. 1.581
37. Summative assessments are more important than formative assessments. 1.285
17. Assessment becomes more accurate due to the longitudinal relationship between supervisor and trainee. 1.284
45. Previous experiences with this trainee influence my assessment. 1.224
46. My style of giving feedback is influenced by the way I expect it to be received. 1.153

Factor 5 $1 > 5$

14. Assessment should primarily drive trainees' learning process. 2.346
38. Professional tasks can be entrusted earlier to a trainee who self-directs his or her learning process. 2.271
34. A trainee's perspectives on the profession of general practice affect his/her assessment. 1.503
12. It is important for a trainee to ask feedback. 1.181
23. Clear and precise assessment criteria are needed to assess a trainee accurately. 1.029

 $5 > 1$

1. Assessment of competency development in General Practice may only take place in the workplace setting. 2.209
15. Competencies cannot be evaluated with (numerical) grades. 1.887
39. For the progressive development of competencies a trainees' learning goals are more important than formal assessment criteria. 1.362
26. An experienced supervisor is capable of conducting more accurate assessments. 1.210
9. Assessment practices assure high quality patient care by the trainee. 1.157

 Factor 2 Descending array of differences

Factor 1 $2 > 1$

2. Acceptance of negative feedback necessitates a relationship of trust between trainee and supervisor. 1.854
35. Assessors should judge in an identical fashion. 1.680
8. Learners should be able to compensate for poor grades over time. 1.603
25. When conducting an assessment, progressive development is more important than actual performance. 1.500
39. For the progressive development of competencies a trainees' learning goals are more important than formal assessment criteria. 1.446

 $1 > 2$

9. Assessment practices assure high quality patient care by the trainee 2.119
23. Clear and precise assessment criteria are needed to assess a trainee accurately. 1.748
38. Professional tasks can be entrusted earlier to a trainee who self-directs his or her learning process. 1.602
14. Assessment should primarily drive trainees' learning process. 1.571
27. As an assessor I feel involved with the training institute. 1.436
-

Factor 2	Descending array of differences	
<i>Factor 3</i>	<p>2 > 3</p> <p>10. Assessment should be based on the trainees' learning goals and, consequently, be tailored to the individual trainee. 2.304</p> <p>45. Previous experiences with this trainee influence my assessment. 2.163</p> <p>17. Assessment becomes more accurate due to the longitudinal relationship between supervisor and trainee. 2.119</p> <p>2. Acceptance of negative feedback necessitates a relationship of trust between trainee and supervisor. 1.887</p> <p>8. Learners should be able to compensate for poor grades over time. 1.751</p>	<p>3 > 2</p> <p>23. Clear and precise assessment criteria are needed to assess a trainee accurately. 2.392</p> <p>27. As an assessor I feel involved with the training institute. 2.212</p> <p>32. Assessment practices stimulate the competency development of trainees. 1.803</p> <p>16. A constructive cooperation between GP supervisor and trainee interferes with critical assessment practices. 1.784</p> <p>24. Numerical grades are not suitable for formative assessments. 1.729</p>
<i>Factor 4</i>	<p>2 > 4</p> <p>25. When conducting an assessment, progressive development is more important than actual performance. 2.186</p> <p>18. Professional tasks are more easily entrusted to a trainee whose range of ideas and practices are similar to those of the GP supervisor. 1.970</p> <p>42. Rigorous assessment requires that both trainee and supervisor can receive feedback. 1.939</p> <p>8. Learners should be able to compensate for poor grades over time. 1.844</p> <p>33. It is important that trainee and GP supervisor have shared perspectives on the GP profession. 1.659</p>	<p>4 > 2</p> <p>27. As an assessor I feel involved with the training institute. 2.478</p> <p>46. My style of giving feedback is influenced by the way I expect it to be received. 2.024</p> <p>7. As an assessor, I feel appreciated by the training institute. 1.936</p> <p>34. A trainee's perspectives on the profession of general practice affect his/her assessment. 1.404</p> <p>11. Knowing whether an assessment is formative or summative is important. 1.244</p>
<i>Factor 5</i>	<p>2 > 5</p> <p>8. Learners should be able to compensate for poor grades over time. 2.164</p> <p>25. When conducting an assessment, progressive development is more important than actual performance. 1.844</p> <p>2. Acceptance of negative feedback necessitates a relationship of trust between trainee and supervisor. 1.813</p> <p>10. Assessment should be based on the trainees' learning goals and, consequently, be tailored to the individual trainee. 1.689</p> <p>35. Assessors should judge in an identical fashion. 1.659</p>	<p>5 > 2</p> <p>9. Assessment practices assure high quality patient care by the trainee. 3.276</p> <p>27. As an assessor I feel involved with the training institute. 2.363</p> <p>7. As an assessor, I feel appreciated by the training institute. 2.053</p> <p>32. Assessment practices stimulate the competency development of trainees. 1.575</p> <p>28. Assessment implies an additional workload. 1.380</p>

Factor 3 Descending array of differences

Factor 1	3 > 1	1 > 3
	15. <i>Competencies cannot be evaluated with (numerical) grades</i> 1.728	46. <i>My style of giving feedback is influenced by the way I expect it to be received.</i> 2.066
	24. <i>Numerical grades are not suitable for formative assessments.</i> 1.611	45. <i>Previous experiences with this trainee influence my assessment.</i> 1.971
	28. <i>Assessment implies an additional workload.</i> 1.349	12. <i>It is important for a trainee to ask feedback.</i> 1.734
	35. <i>Assessors should judge in an identical fashion.</i> 1.278	38. <i>Professional tasks can be entrusted earlier to a trainee who self-directs his or her learning process.</i> 1.465
	48. <i>It is important to document assessments regularly.</i> 1.258	10. <i>Assessment should be based on the trainees' learning goals and, consequently, be tailored to the individual trainee.</i> 1.153
Factor 2	3 > 2	2 > 3
	23. <i>Clear and precise assessment criteria are needed to assess a trainee accurately.</i> 2.392	10. <i>Assessment should be based on the trainees' learning goals and, consequently, be tailored to the individual trainee.</i> 2.304
	27. <i>As an assessor I feel involved with the training institute.</i> 2.212	45. <i>Previous experiences with this trainee influence my assessment.</i> 2.163
	32. <i>Assessment practices stimulate the competency development of trainees.</i> 1.803	17. <i>Assessment becomes more accurate due to the longitudinal relationship between supervisor and trainee.</i> 2.119
	16. <i>A constructive cooperation between GP supervisor and trainee interferes with critical assessment practices.</i> 1.784	2. <i>Acceptance of negative feedback necessitates a relationship of trust between trainee and supervisor.</i> 1.887
	24. <i>Numerical grades are not suitable for formative assessments.</i> 1.729	8. <i>Learners should be able to compensate for poor grades over time.</i> 1.751
Factor 4	3 > 4	4 > 3
	23. <i>Clear and precise assessment criteria are needed to assess a trainee accurately.</i> 1.989	46. <i>My style of giving feedback is influenced by the way I expect it to be received.</i> 3.219
	14. <i>Assessment should primarily drive trainees' learning process</i> 1.766	45. <i>Previous experiences with this trainee influence my assessment.</i> 3.195
	24. <i>Numerical grades are not suitable for formative assessments.</i> 1.723	17. <i>Assessment becomes more accurate due to the longitudinal relationship between supervisor and trainee.</i> 1.997
	15. <i>Competencies cannot be evaluated with (numerical) grades.</i> 1.435	2. <i>Acceptance of negative feedback necessitates a relationship of trust between trainee and supervisor.</i> 1.614
	32. <i>Assessment practices stimulate the competency development of trainees.</i> 1.347	20. <i>A capable trainee is easy to recognize.</i> 1.160
Factor 5	3 > 5	5 > 3
	23. <i>Clear and precise assessment criteria are needed to assess a trainee accurately.</i> 1.674	45. <i>Previous experiences with this trainee influence my assessment.</i> 2.416
	14. <i>Assessment should primarily drive trainees' learning process.</i> 1.616	9. <i>Assessment practices assure high quality patient care by the trainee.</i> 2.246
	16. <i>A constructive cooperation between GP supervisor and trainee interferes with critical assessment practices.</i> 1.473	46. <i>My style of giving feedback is influenced by the way I expect it to be received.</i> 1.849
	35. <i>Assessors should judge in an identical fashion.</i> 1.257	1. <i>Assessment of competency development in General Practice may only take place in the workplace setting.</i> 1.342
	48. <i>It is important to document assessments regularly.</i> 1.200	39. <i>For the progressive development of competencies a trainees' learning goals are more important than formal assessment criteria.</i> 1.244

Factor 4 Descending array of differences

<p>Factor 1</p> <p>4 > 1</p> <p>2. <i>Acceptance of negative feedback necessitates a relationship of trust between trainee and supervisor.</i> 1.581</p> <p>37. <i>Summative assessments are more important than formative assessments.</i> 1.285</p> <p>17. <i>Assessment becomes more accurate due to the longitudinal relationship between supervisor and trainee.</i> 1.284</p> <p>45. <i>Previous experiences with this trainee influence my assessment.</i> 1.224</p> <p>46. <i>My style of giving feedback is influenced by the way I expect it to be ask feedback.</i> 1.153</p> <p>Factor 2</p> <p>4 > 2</p> <p>27. <i>As an assessor I feel involved with the training institute.</i> 2.478</p> <p>46. <i>My style of giving feedback is influenced by the way I expect it to be received.</i> 2.024</p> <p>7. <i>As an assessor, I feel appreciated by the training institute.</i> 1.936</p> <p>34. <i>A trainee's perspectives on the profession of general practice affect his/her assessment.</i> 1.404</p> <p>11. <i>Knowing whether an assessment is formative or summative is important.</i> 1.244</p> <p>2</p> <p>Factor 3</p> <p>4 > 3</p> <p>46. <i>My style of giving feedback is influenced by the way I expect it to be received.</i> 3.219</p> <p>45. <i>Previous experiences with this trainee influence my assessment.</i> 3.195</p> <p>17. <i>Assessment becomes more accurate due to the longitudinal relationship between supervisor and trainee.</i> 1.997</p> <p>2. <i>Acceptance of negative feedback necessitates a relationship of trust between trainee and supervisor.</i> 1.614</p> <p>20. <i>A capable trainee is easy to recognize.</i> 1.160</p>	<p>1 > 4</p> <p>14. <i>Assessment should primarily drive trainees' learning process.</i> 2.495</p> <p>18. <i>Professional tasks are more easily entrusted to a trainee whose range of ideas and practices are similar to those of the GP supervisor.</i> 1.941</p> <p>40. <i>When assessing a trainee, it is crucial that a trainee can perform professional tasks independently.</i> 1.631</p> <p>23. <i>Clear and precise assessment criteria are needed to assess a trainee accurately.</i> 1.345</p> <p>12. <i>It is important for a trainee to ask feedback.</i> 1.299</p> <p>2 > 4</p> <p>25. <i>When conducting an assessment, progressive development is more important than actual performance.</i> 2.186</p> <p>18. <i>Professional tasks are more easily entrusted to a trainee whose range of ideas and practices are similar to those of the GP supervisor.</i> 1.970</p> <p>42. <i>Rigorous assessment requires that both trainee and supervisor can receive feedback.</i> 1.939</p> <p>8. <i>Learners should be able to compensate for poor grades over time.</i> 1.844</p> <p>33. <i>It is important that trainee and GP supervisor have shared perspectives on the GP profession.</i> 1.659</p> <p>3 > 4</p> <p>23. <i>Clear and precise assessment criteria are needed to assess a trainee accurately.</i> 1.989</p> <p>14. <i>Assessment should primarily drive trainees' learning process</i> 1.766</p> <p>24. <i>Numerical grades are not suitable for formative assessments.</i> 1.723</p> <p>15. <i>Competencies cannot be evaluated with (numerical) grades.</i> 1.435</p> <p>32. <i>Assessment practices stimulate the competency development of trainees.</i> 1.347</p>
---	--

 Factor 4 Descending array of differences

Factor 5 4 > 5

38. Professional tasks can be entrusted earlier to a trainee who self-directs his or her learning process. 1.785
2. Acceptance of negative feedback necessitates a relationship of trust between trainee and supervisor. 1.540
34. Perspectives of a trainee on the profession of General Practice affect his assessment. 1.484
46. My style of giving feedback is influenced by the way I expect it to be received. 1.370
17. Assessment becomes more accurate due to the longitudinal relationship between GP supervisor and trainee. 1.308

5 > 4

1. Assessment of competency development in General Practice may only take place in the workplace setting. 2.404
9. Assessment practices assure high quality patient care by the trainee. 2.277
33. It is important that trainee and GP supervisor have shared perspectives on the GP profession. 1.802
42. Rigorous assessment requires that both trainee and supervisor can receive feedback. 1.778
40. When assessing a trainee, it is crucial that a trainee can perform professional tasks independently. 1.600

 Factor 5 Descending array of differences

Factor 1 5 > 1

1. Assessment of competency development in General Practice may only take place in the workplace setting. 2.209
15. Competencies cannot be evaluated with (numerical) grades. 1.887
39. For the progressive development of competencies a trainees' learning goals are more important than formal assessment criteria. 1.362
26. An experienced supervisor is capable of conducting more accurate assessments. 1.210
9. Assessment practices assure high quality patient care by the trainee. 1.157

1 > 5

14. Assessment should primarily drive trainees' learning process. 2.346
38. Professional tasks can be entrusted earlier to a trainee who self-directs his or her learning process. 2.271
34. A trainee's perspectives on the profession of general practice affect his/her assessment. 1.503
12. It is important for a trainee to ask feedback. 1.181
23. Clear and precise assessment criteria are needed to assess a trainee accurately. 1.029

Factor 2 5 > 2

9. Assessment practices assure high quality patient care by the trainee. 3.276
27. As an assessor I feel involved with the training institute. 2.363
7. As an assessor, I feel appreciated by the training institute. 2.053
32. Assessment practices stimulate the competency development of trainees. 1.575
28. Assessment implies an additional workload. 1.380

2 > 5

8. Learners should be able to compensate for poor grades over time. 2.164
25. When conducting an assessment, progressive development is more important than actual performance. 1.844
2. Acceptance of negative feedback necessitates a relationship of trust between trainee and supervisor. 1.813
10. Assessment should be based on the trainees' learning goals and, consequently, be tailored to the individual trainee. 1.689
35. Assessors should judge in an identical fashion. 1.659
-

Factor 5	Descending array of differences	
<i>Factor 3</i>	<p>5 > 3</p> <p>45. Previous experiences with this trainee influence my assessment. 2.416</p> <p>9. Assessment practices assure high quality patient care by the trainee. 2.246</p> <p>46. My style of giving feedback is influenced by the way I expect it to be received. 1.849</p> <p>1. Assessment of competency development in General Practice may only take place in the workplace setting. 1.342</p> <p>39. For the progressive development of competencies a trainees' learning goals are more important than formal assessment criteria. 1.244</p>	<p>3 > 5</p> <p>23. Clear and precise assessment criteria are needed to assess a trainee accurately. 1.674</p> <p>14. Assessment should primarily drive trainees' learning process. 1.616</p> <p>16. A constructive cooperation between GP supervisor and trainee interferes with critical assessment practices. 1.473</p> <p>35. Assessors should judge in an identical fashion. 1.257</p> <p>48. It is important to document assessments regularly. 1.200</p>
<i>Factor 4</i>	<p>5 > 4</p> <p>1. Assessment of competency development in General Practice may only take place in the workplace setting. 2.404</p> <p>9. Assessment practices assure high quality patient care by the trainee. 2.277</p> <p>33. It is important that trainee and GP supervisor have shared perspectives on the GP profession. 1.802</p> <p>42. Rigorous assessment requires that both trainee and supervisor can receive feedback. 1.778</p> <p>40. When assessing a trainee, it is crucial that a trainee can perform professional tasks independently. 1.600</p>	<p>4 > 5</p> <p>38. Professional tasks can be entrusted earlier to a trainee who self-directs his or her learning process. 1.785</p> <p>2. Acceptance of negative feedback necessitates a relationship of trust between trainee and supervisor. 1.540</p> <p>34. A trainee's perspectives on the profession of general practice affect his/her assessment. 1.484</p> <p>46. My style of giving feedback is influenced by the way I expect it to be received. 1.370</p> <p>17. Assessment becomes more accurate due to the longitudinal relationship between supervisor and trainee. 1.308</p>

^a The descending array of differences shows the differences between Z-scores of any pair of factors

References

- Ajzen, I. (2002). Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior. *Journal of Applied Social Psychology*, 32(4), 665–683. doi:10.1111/j.1559-1816.2002.tb00236.x.
- Ajzen, I., & Madden, T. J. (1986). Prediction of goal-directed behavior: Attitudes, intentions, and perceived behavioral control. *Journal of Experimental Social Psychology*, 22(5), 453–474. doi:10.1016/0022-1031(86)90045-4.
- Archer, J. C. (2010). State of the science in health professional education: Effective feedback. *Medical Education*, 44(1), 101–108. doi:10.1111/j.1365-2923.2009.03546.x.
- Berendonk, C., Stalmeijer, R. E., & Schuwirth, L. W. (2013). Expertise in performance assessment: assessors' perspectives. *Advances in health sciences education : theory and practice*, 18(4), 559–571. doi:10.1007/s10459-012-9392-x.
- Bok, H. G., Teunissen, P. W., Favier, R. P., Rietbroek, N. J., Theyse, L. F., Brommer, H., et al. (2013). Programmatic assessment of competency-based workplace learning: When theory meets practice. *BMC Medical Education*, 13(1), 123. doi:10.1186/1472-6920-13-123.
- Brown, S. R. (2009). Q technique, method, and methodology: Comments on Stentor Danielson's article. *Field Methods*, 21(3), 238–241. doi:10.1177/1525822x09332080.
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning—A theoretical synthesis. *Review of Educational Research*, 65(3), 245–281. doi:10.3102/00346543065003245.
- Cilliers, F., & Schuwirth, L. (2015). *Health behaviour theories: A conceptual lens to explore behaviour change researching medical education* (pp. 141–154). Hoboken: Wiley.

- Cook, D. A., Beckman, T. J., Mandrekar, J. N., & Pankratz, V. S. (2010). Internal structure of mini-CEX scores for internal medicine residents: Factor analysis and generalizability. *Advances Health Sciences Education Theory and Practice*, 15(5), 633–645. doi:[10.1007/s10459-010-9224-9](https://doi.org/10.1007/s10459-010-9224-9).
- Cook, D. A., Dupras, D. M., Beckman, T. J., Thomas, K. G., & Pankratz, V. S. (2009). Effect of rater training on reliability and accuracy of mini-cex scores: A randomized, controlled trial. *Journal of General Internal Medicine*, 24(1), 74–79. doi:[10.1007/s11606-008-0842-3](https://doi.org/10.1007/s11606-008-0842-3).
- Cross, R. M. (2005). Exploring attitudes: The case for Q methodology. *Health Education Research*, 20(2), 206–213. doi:[10.1093/Her/Cyg121](https://doi.org/10.1093/Her/Cyg121).
- Di Genova, T., Valentino, P. L., Gosselin, R., & Bhanji, F. (2015). The academic half-day redesigned: Improving generalism, promoting CanMEDS and developing self-directed learners. *Paediatrics and Child Health*, 20(1), 30–34.
- Embo, M. P., Driessen, E. W., Valcke, M., & Van der Vleuten, C. P. (2010). Assessment and feedback to facilitate self-directed learning in clinical practice of Midwifery students. *Medical Teacher*, 32(7), e263–e269. doi:[10.3109/0142159X.2010.490281](https://doi.org/10.3109/0142159X.2010.490281).
- Eva, K. W., & Hodges, B. D. (2012). Scylla or Charybdis? Can we navigate between objectification and judgement in assessment? *Medical Education*, 46(9), 914–919. doi:[10.1111/j.1365-2923.2012.04310.x](https://doi.org/10.1111/j.1365-2923.2012.04310.x).
- Fokkema, J. P., Scheele, F., Westerman, M., van Exel, J., Scherpier, A. J., van der Vleuten, C. P., et al. (2014). Perceived effects of innovations in postgraduate medical education: A Q study focusing on workplace-based assessment. *Academic Medicine*. doi:[10.1097/ACM.0000000000000394](https://doi.org/10.1097/ACM.0000000000000394).
- Frank, J. R., Snell, L. S., Sherbino, J., editors. Draft CanMEDS. (2015). Physician Competency Framework – Series III. Ottawa: The Royal College of Physicians and Surgeons of Canada. doi:[10.13140/2.1.5039.4402](https://doi.org/10.13140/2.1.5039.4402).
- Gawronski, B., Ehrenberg, K., Banse, R., Zukova, J., & Klauer, K. C. (2003). It's in the mind of the beholder: The impact of stereotypic associations on category-based and individuating impression formation. *Journal of Experimental Social Psychology*, 39(1), 16–30. doi:[10.1016/S0022-1031\(02\)00517-6](https://doi.org/10.1016/S0022-1031(02)00517-6).
- Gingerich, A., Kogan, J., Yeates, P., Govaerts, M., & Holmboe, E. (2014). Seeing the 'black box' differently: assessor cognition from three research perspectives. *Medical Education*, 48(11), 1055–1068. doi:[10.1111/medu.12546](https://doi.org/10.1111/medu.12546).
- Gingerich, A., Regehr, G., & Eva, K. W. (2011). Rater-based assessments as social judgments: Rethinking the etiology of rater errors. *Academic Medicine : Journal of the Association of American Medical Colleges*, 86(10 Suppl), S1–S7. doi:[10.1097/ACM.0b013e31822a6cf8](https://doi.org/10.1097/ACM.0b013e31822a6cf8).
- Govaerts, M. J., van de Wiel, M. W., & van der Vleuten, C. P. (2013). Quality of feedback following performance assessments: Does assessor expertise matter? *European Journal of Training and Development*, 37(1), 105–125. doi:[10.1108/03090591311293310](https://doi.org/10.1108/03090591311293310).
- Govaerts, M., & van der Vleuten, C. P. M. (2013). Validity in work-based assessment: Expanding our horizons. *Medical Education*, 47(12), 1164–1174. doi:[10.1111/medu.12289](https://doi.org/10.1111/medu.12289).
- Govaerts, M. J., van der Vleuten, C. P., Schuwirth, L. W., & Muijtens, A. M. (2007). Broadening perspectives on clinical performance assessment: Rethinking the nature of in-training assessment. *Advances in Health Sciences Education : Theory and Practice*, 12(2), 239–260. doi:[10.1007/s10459-006-9043-1](https://doi.org/10.1007/s10459-006-9043-1).
- Ha, E. H. (2014). Attitudes toward video-assisted debriefing after simulation in undergraduate nursing students: an application of Q methodology. *Nurse Education Today*, 34(6), 978–984. doi:[10.1016/j.nedt.2014.01.003](https://doi.org/10.1016/j.nedt.2014.01.003).
- Harasym, P. H., Woloschuk, W., & Cuning, L. (2008). Undesired variance due to examiner stringency/leniency effect in communication skill scores assessed in OSCEs. *Advances Health Sciences Education Theory and Practice*, 13(5), 617–632. doi:[10.1007/s10459-007-9068-0](https://doi.org/10.1007/s10459-007-9068-0).
- Harvey, H., Good, J., Mason, J., & Reissland, N. (2013). A Q-methodology study of parental understandings of infant immunisation: Implications for health-care advice. *Journal of Health Psychology*. doi:[10.1177/1359105313513622](https://doi.org/10.1177/1359105313513622).
- Hawkins, R. E., Welcher, C. M., Holmboe, E. S., Kirk, L. M., Norcini, J. J., Simons, K. B., et al. (2015). Implementation of competency-based medical education: Are we addressing the concerns and challenges? *Medical Education*, 49(11), 1086–1102. doi:[10.1111/medu.12831](https://doi.org/10.1111/medu.12831).
- Hodges, B. (2013). Assessment in the post-psychometric era: Learning to love the subjective and collective. *Medical Teacher*, 35(7), 564–568. doi:[10.3109/0142159X.2013.789134](https://doi.org/10.3109/0142159X.2013.789134).
- Hodges, B., Regehr, G., & Martin, D. (2001). Difficulties in recognizing one's own incompetence: Novice physicians who are unskilled and unaware of it. *Academic Medicine*, 76(10 Suppl), S87–S89.
- Holmboe, E. S. (2014). Realizing the promise of competency-based medical education. *Academic Medicine*. doi:[10.1097/ACM.0000000000000515](https://doi.org/10.1097/ACM.0000000000000515).

- Holmboe, E. S., Hawkins, R. E., & Huot, S. J. (2004). Effects of training in direct observation of medical residents' clinical competence—A randomized trial. *Annals of Internal Medicine*, *140*(11), 874–881.
- Holmboe, E. S., Sherbino, J., Long, D. M., Swing, S. R., & Frank, J. R. (2010). The role of assessment in competency-based medical education. *Medical Teacher*, *32*(8), 676–682. doi:10.3109/0142159X.2010.500704.
- Holmboe, E. S., Ward, D. S., Reznick, R. K., Katsufakis, P. J., Leslie, K. M., Patel, V. L., et al. (2011). Faculty development in assessment: the missing link in competency-based medical education. *Academic Medicine : Journal of the Association of American Medical Colleges*, *86*(4), 460–467. doi:10.1097/ACM.0b013e31820cb2a7.
- Honey, S., Bryant, L. D., Murray, J., Hill, K., & House, A. (2013). Differences in the perceived role of the healthcare provider in delivering vascular health checks: A Q methodology study. *BMC family practice*, *14*, 172. doi:10.1186/1471-2296-14-172.
- Kogan, J. R., Conforti, L., Bernabeo, E., Iobst, W., & Holmboe, E. (2011). Opening the black box of clinical skills assessment via observation: A conceptual model. *Medical Education*, *45*(10), 1048–1060. doi:10.1111/j.1365-2923.2011.04025.x.
- Kogan, J., Conforti, L. N., Bernabeo, E., Iobst, W., & Holmboe, E. (2015). How faculty members experience workplace-based assessment rater training: A qualitative study. *Medical Education*, *49*(7), 692–708. doi:10.1111/medu.12733.
- Massie, J., & Ali, J. M. (2015). Workplace-based assessment: A review of user perceptions and strategies to address the identified shortcomings. *Advances in Health Sciences Education*, *21*(2), 455–473. doi:10.1007/s10459-015-9614-0.
- McManus, I. C., Thompson, M., & Mollon, J. (2006). Assessment of examiner leniency and stringency ('hawk-dove effect') in the MRCP(UK) clinical examination (PACES) using multi-facet Rasch modelling. *BMC Medical Education*, *6*, 42. doi:10.1186/1472-6920-6-42.
- Meade, L. B., Caverzagie, K. J., Swing, S. R., Jones, R. R., O'Malley, C. W., Yamazaki, K., et al. (2013). Playing with curricular milestones in the educational sandbox: Q-sort results from an internal medicine educational collaborative. *Academic Medicine*, *88*(8), 1142–1148. doi:10.1097/ACM.0b013e31829a3967.
- Milcu, A. I., Sherren, K., Hanspach, J., Abson, D., & Fischer, J. (2014). Navigating conflicting landscape aspirations: Application of a photo-based Q-method in Transylvania (Central Romania). *Land Use Policy*, *41*, 408–422. doi:10.1016/j.landusepol.2014.06.019.
- Murphy, D. J., Bruce, D., & Eva, K. W. (2008). Workplace-based assessment for general practitioners: Using stakeholder perception to aid blueprinting of an assessment battery. *Medical Education*, *42*(1), 96–103. doi:10.1111/j.1365-2923.2007.02952.x.
- Newell, B. R., & Shanks, D. R. (2014). Unconscious influences on decision making: A critical review. *The Behavioral and Brain Sciences*, *37*(1), 1–19. doi:10.1017/S0140525X12003214.
- Norcini, J., & Burch, V. (2007). Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Medical Teacher*, *29*(9), 855–871. doi:10.1080/01421590701775453.
- Pelgrim, E. A., Kramer, A. W., Morkink, H. G., & van der Vleuten, C. P. (2012). The process of feedback in workplace-based assessment: Organisation, delivery, continuity. *Medical Education*, *46*(6), 604–612. doi:10.1111/j.1365-2923.2012.04266.x.
- Regehr, G., & Eva, K. (2006). Self-assessment, self-direction, and the self-regulating professional. *Clinical Orthopaedics and Related Research*, *449*, 34–38. doi:10.1097/01.blo.0000224027.85732.b2.
- Sackett, P. R., Laczko, R. M., & Arvey, R. D. (2002). The effects of range restriction on estimates of criterion interrater reliability: Implications for validation research. *Personnel Psychology*, *55*(4), 807–825. doi:10.1111/j.1744-6570.2002.tb00130.x.
- Sandars, J., & Cleary, T. J. (2011). Self-regulation theory: Applications to medical education: AMEE Guide No. 58. *Medical Teacher*, *33*(11), 875–886. doi:10.3109/0142159X.2011.595434.
- Schmolck, P. (2014). PQMethod software. <http://schmolck.userweb.mwn.de/qmethod/>.
- Schoenherr, J. R., & Hamstra, S. J. (2016). Psychometrics and its discontents: An historical perspective on the discourse of the measurement tradition. *Advances Health Sciences Education Theory and Practices*, *21*(3), 719–729. doi:10.1007/s10459-015-9623-z.
- Schuwirth, L., & Swanson, D. (2013). Standardised versus individualised assessment: Related problems divided by a common language. *Medical Education*, *47*(6), 627–631. doi:10.1111/medu.12111.
- Shabila, N. P., Al-Tawil, N. G., Al-Hadithi, T. S., & Sondorp, E. (2014). Using Q-methodology to explore people's health seeking behavior and perception of the quality of primary care services. *BMC public health*, *14*, 2. doi:10.1186/1471-2458-14-2.
- Shay, S. B. (2004). The assessment of complex performance: A socially situated interpretive act. *Harvard Educational Review*, *74*(3), 307–329.

- Shinebourne, P. (2009). Using Q method in qualitative research. *International Journal of Qualitative Methods*, 8(1), 93–97.
- Stenner, P. H., Cooper, D., & Skevington, S. M. (2003). Putting the Q into quality of life; the identification of subjective constructions of health-related quality of life using Q methodology. *Social Science and Medicine*, 57(11), 2161–2172.
- Stenner, P. H., Dancey, C. P., & Watts, S. (2000). The understanding of their illness amongst people with irritable bowel syndrome: A Q methodological study. *Social Science and Medicine*, 51(3), 439–452.
- ten Cate, O., & Scheele, F. (2007). Competency-based postgraduate training: Can we bridge the gap between theory and clinical practice? *Academic Medicine : Journal of the Association of American Medical Colleges*, 82(6), 542–547. doi:10.1097/ACM.0b013e31805559c7.
- Teunissen, P. W., & Bok, H. G. (2013). Believing is seeing: how people's beliefs influence goals, emotions and behaviour. *Medical Education*, 47(11), 1064–1072. doi:10.1111/medu.12228.
- Tochel, C., Haig, A., Heskeith, A., Cadzow, A., Beggs, K., Colthart, I., et al. (2009). The effectiveness of portfolios for post-graduate assessment and education: BEME Guide No 12. *Medical Teacher*, 31(4), 299–318. doi:10.1080/01421590902883056.
- Tziner, A., Murphy, K. R., & Cleveland, J. N. (2001). Relationships between attitudes toward organizations and performance appraisal systems and rating behavior. *International Journal of Selection and Assessment*, 9(3), 226–239. doi:10.1111/1468-2389.00176.
- Tziner, A., Murphy, K. R., & Cleveland, J. N. (2005). Contextual and rater factors affecting rating behavior. *Group and Organization Management*, 30(1), 89–98. doi:10.1177/1059601104267920.
- Tziner, A., Murphy, K. R., Cleveland, J. N., Beaudin, G., & Marchand, S. (1998). Impact of rater beliefs regarding performance appraisal and its organizational context on appraisal quality. *Journal of Business and Psychology*, 12(4), 457–467. doi:10.1023/A:1025003106150.
- van der Vleuten, C. P., Schuwirth, L. W., Driessen, E. W., Govaerts, M. J., & Heeneman, S. (2014). Twelve tips for programmatic assessment. *Medical Teacher*. doi:10.3109/0142159X.2014.973388.
- van der Vleuten, C. P., Schuwirth, L. W., Scheele, F., Driessen, E. W., & Hodges, B. (2010). The assessment of professional competence: Building blocks for theory development. *Best Practice & Research. Clinical Obstetrics & Gynaecology*. doi:10.1016/j.bpobgyn.2010.04.001.
- van der Vleuten, C., & Verhoeven, B. H. (2013). In-training assessment developments in postgraduate education in Europe. *ANZ Journal of Surgery*, 83(6), 454–459. doi:10.1111/ans.12190.
- Van Exel, N. J. A. (2005). *Q methodology: A sneak preview*. [Available from www.jobvanexel.nl].
- Wallenburg, I., van Exel, J., Stolk, E., Scheele, F., de Bont, A., & Meurs, P. (2010). Between trust and accountability: Different perspectives on the modernization of postgraduate medical training in the Netherlands. *Academic Medicine : Journal of the Association of American Medical Colleges*, 85(6), 1082–1090. doi:10.1097/ACM.0b013e3181dc1f0f.
- Watling, C. J., & Lingard, L. (2012). Toward meaningful evaluation of medical trainees: The influence of participants' perceptions of the process. *Advances in Health Sciences Education : Theory and Practice*, 17(2), 183–194. doi:10.1007/s10459-010-9223-x.
- Watts, S., & Stenner, P. (2005a). Doing Q methodology: theory, method and interpretation. *Qualitative Research in Psychology*, 2(1), 67–91. doi:10.1191/1478088705qp0220a.
- Watts, S., & Stenner, P. (2005b). The subjective experience of partnership love: A Q Methodological study. *The British Journal of Social Psychology/The British Psychological Society*, 44(Pt 1), 85–107. doi:10.1348/014466604X23473.
- Watts, S., & Stenner, P. (2012). *Doing Q methodological research: Theory, method and interpretation*. London: SAGE.
- Yeates, P., O'Neill, P., Mann, K., & Eva, K. W. (2013). 'You're certainly relatively competent': Assessor bias due to recent experiences. *Medical Education*, 47(9), 910–922. doi:10.1111/medu.12254.