ORIGINAL RESEARCH

Performance standards for teachers supporting nursing students' reflection skills development

Agaath Dekker Groen, Marieke van der Schaaf, Karel Stokking

Department of Educational Sciences, Utrecht University, Utrecht, the Netherlands

Correspondence: A.M. Dekker-Groen. Address: PO-Box 80140, 3508 TC Utrecht, the Netherlands. Telephone: 31-30-253-4700. E-mail: a.m.dekker@uu.nl

Received: October 19, 2011 **Accepted:** November 29, 2011 **Published:** February 1, 2012
DOI: 10.5430/jnep.v2n1p9 **URL:** http://dx.doi.org/10.5430/jnep.v2n1p9

Abstract

In the study performance standards for teachers were developed and validated. A ten-step procedure was followed to ensure procedural and internal validity. National competences and specific content standards for supporting nursing reflection skills development formed the foundation of a preliminary rubric framework which was piloted. Forty participants from six nursing institutes judged the developed rubric framework of eight competences covering thirty rubric attributes. They also discussed the prerequisite minimum performance level and judgmental models. These judgments and discussions resulted in consensus on the rubric framework, a general cut-off score, and a conjunctive judgmental model that is convenient for assessing nursing teachers' competences. The rubrics can be used in a teacher training program. Also institutes of nursing education can employ the rubrics as a tool for preparing and formatively assessing students' reflection skills.

Key words

Nursing education, Performance standards, Reflection skills, Rubrics, Teacher competences

1 Introduction

Health care professionals and teachers are continuously facing changes to working practices and technologies. To respond to these changes they are given new roles and need to develop their competences [1, 2]. Consequently, students as future professionals have to be equipped with skills for life-long learning, e.g. reflection on own work and competences [3-5]. Reflection skills are difficult to develop [6] and students need support for acquiring and improving these skills [7]. Coaching and constructive feedback turn out to be essential for acquiring reflection skills [8]. Feedback is information communicated to students for providing changes in thinking or performance [9]. Since feedback is information used to close the gap between current performance and certain standards, it should be based on those standards [10, 11]. Similarly to students their teachers also need feedback to acquire and develop competences to support the development of students' reflection skills. This feedback has to inform teachers about the desired performances they should carry out. International attention for monitoring teacher performances and improving teacher quality has led to the development of teacher standards [12, 13]. However, such standards are not yet specified regarding supporting students' reflection.

The study was conducted in the Netherlands in nursing education at the level of senior secondary vocational education (level IV of the International Standard Classification of Education, UNESCO). The goal of senior secondary vocational

education is to deliver students to the labor market as well as to prepare and stimulate students to continue their educational career in higher professional education.

2 Background

2.1 Standard-setting

Teacher competences such as their integrated knowledge, skills and attitudes, are needed to perform professional tasks. Although some competences cannot be observed directly, teacher behavior can be observed. Standards for teacher competences include content standards and performance standards. Content standards specify what teachers are expected to know and be able to do [14]. Performance standards specify norms for the frequency, intensity or quality on which these performance aspects will be scored while assessing teachers [15, 16] or other professionals in the workplace [17]. Currently performance standards are increasingly based on judgments of a group of selected experts about performance levels and cut-off scores. A cut-off score indicates a minimum required or "just sufficient" performance level acceptable to judges [14, 18, 19].

A convenient standard-setting method for teachers supporting nursing students' reflection skills is using rubrics to describe and judge standards. A rubric consists of parts or aspects of work that constitute its quality together with several performance level descriptions regarding each particular task [20, 21].

2.2 Judgmental model

In setting performance standards not only a standard-setting method should be chosen but also a judgmental model. A judgmental model is used to come to a final judgment about the performance based on scores for each task on each standard. To combine these scores, several judgmental models can be applied: a compensatory model, a conjunctive model, or a disjunctive model [14, 18]. In a compensatory model the total score obtained on the different standards determines whether the overall performance is sufficient. In this model high scores on certain standards can compensate for low scores on other standards. In a conjunctive model a certain minimal score has to be reached on each standard. In a disjunctive model for certain standards a minimal score is mandatory and for other standards low scores are allowed if high scores on other standards compensate enough to reach the minimum total score.

2.3 Validation strategies

Rubrics have to meet certain quality requirements. In validation strategies different kinds of evidence play a role: procedural, internal, and external validity evidence [19]. Procedural evidence consists of demonstrating the soundness of a procedure to develop rubrics. To warrant procedural validity the standard-setting procedure has to be clearly and explicitly defined and documented. The implementation should be easy and systematic. Moreover, judges have to know the purpose for which the standards are being set and understand and feel comfortable with the standard setting procedure they are using [14]. Internal validation evidence focuses on the consistency of judges in translating ratings on performance standards in a cut-off score. To warrant internal validity, rubrics have to be representative of the target domain. Moreover, rubrics have to be clear to the judges [19], who rate the rubrics independently [18] and reach consensus about the judgments [14]. External evidence can be provided by comparing the resulting performance standards with other sources of information, such as decisions based on performance assessments in authentic situations or in simulations. As teaching and assessing reflection skills is rather new in nursing education and external information is not yet available, this study is restricted to procedural and internal validation strategies. Developing rubrics as well as the standard-setting itself are commonly based on standard-setting procedures that provide procedural and internal evidence.

2.4 Development of the standard-setting procedure

In this study we combined three prominent ways of standard-setting [14, 18, 19]. Berk [18] recommended a generic method for performance standard-setting. Hambleton and Pitoniak [14] explicitly considered the selection of a performance standard setting process and discussed additionally steps which may be essential depending on the method which is chosen. Kane distinguished criteria for the implementation of standard-setting procedures [19]. So the focus of standard setting and the terminology used differ, but there is a considerable overlap. Similar activities are the ones that according to Hambleton and Pitoniak [14], when given careful attention, can substantially increase the defensibility of the performance standards: selection of the method, selection and the training of the judges, consecutive activities from defining standards to computing cut-off scores, and validation and documentation of the total process.

Striving for a procedure as complete and unambiguous as possible we combined these three ways of standard setting and sometimes adapted the terminology for clearness [14, 18, 19]. Based on this combination a standard-setting procedure of ten activities was developed (see Table 1). The article is composed according to these ten activities.

Table 1. Procedure for performance standard-setting

Three ways of performance standard-setting [14, 18, 19] were combined to the standard-setting procedure used in the study. In this procedure ten activities (see below) are distinguished to provide procedural and internal validity evidence. Between square brackets is noticed which source was used.

1. Select a method for the standard-setting procedure [14]
2. Define goals for the standard-setting procedure [19]
3. Prepare performance descriptions: specify types of performance that are expected [14, 18, 19] and the achievement levels [18]
4. Select judges who represent the target group [14, 18, 19]
5. Train the judges: inform them and get them acquainted to the procedure [14, 18, 19]
6. Process of independent judgment: let the participants judge [14, 19] meanwhile providing feedback to the judges and encourage discussions [14, 18] to get consensus [18] on appropriate and clearly stated definitions [19]
7. Compute the cut-off scores based on the judges' ratings and decide about these scores [14, 18]
8. Select a judgmental model [18]
9. Evaluate the standard-setting procedure with judges [14]
10. Check validity [14, 19] and document the findings [14]

2.5 Purpose and research questions

The purpose of the study was the development, validation and setting of performance standards which are relevant and acceptable for nursing teachers to work with in everyday practice to support students' reflection skills development.

The research questions are: What are relevant and acceptable performance standards for teachers who have to support nursing students' development of reflection skills? Which judgmental model is, according to nursing teacher teams, convenient for assessing teachers' competences?

3 Methods

3.1 Development of rubric framework

A rubric method was chosen for standard-setting (activity 1 in Table 1) because rubrics have proven to be convenient for developing performance standards that are feasible for teachers in professional practice of nursing education to support students' reflections (activity 2). Firstly, a rubric framework was developed by integrating national teacher competences and specific content standards for teachers supporting nursing students' reflection skills development. The national teacher competences were chosen because of their legal status in the Netherlands [23]. These competences represent seven general competence domains: (1) interpersonal domain with a focus on effective communication; (2) pedagogical domain to care for a positive learning climate; (3) subject matter and didactics domain concerning instruction, guidance, feedback and assessment of reflection skills; (4) organizational domain to select reflection goals and tasks and use a system for observation, assessment and registration; (5) cooperation within a team about reflection education; (6) cooperation with the environment for example about the students' internship or national curriculum development materials; (7) reflection by and development of the teachers. In our study competence domains 5, 6 and 7 were combined into one competence domain called professional development because these three domains can be seen as a means to professionalism.

Secondly, specific content standards for nursing teachers were developed and validated in an earlier study [24]. These standards are arranged into six task domains. Analyses showed that the first task domain, preparation of reflection instruction, partly corresponds with both the organizational and professional development competence domain. The second task domain, learning goals, links to both the interpersonal and pedagogical competence domain. Further, the other four task domains: instruction, coaching, feedback and assessment correspond with the subject matter and didactics competence domain in which they could form separate competences. Consequently, 72 separate content standards on the different task domains could be used for performance descriptions in the rubric framework while distinguishing eight competences in five competence domains. Per competence two to five rubric attributes were distinguished. For all 30 rubric attributes performance descriptions on three performance levels were developed (activity 3). See Table 2 for the rubric framework.

Table 2. Rubric framework for setting performance standards

Five competence domains	Eight competences	30 rubric attributes*	Content standards 72 of 91 included[24]	α**
Interpersonal	Effective communication with individual and group	4	7	0.88
Pedagogical	Care for a positive learning climate	3	6	0.76
Subject matter and didactical	Give information and instruction how to reflect	5	26	0.95
	Coach the learning process of how to reflect	4	6	0.90
	Give feedback, teach how to receive and ask for feedback	5	11	0.91
	Assess reflections	4	6	0.90
Organizational	Design and organize the learning environment	2	3	0.87
Professional development	Work individually and in cooperation with colleagues	3	7	0.86

* Each rubric attribute got four performance level descriptions.

**Reliability of allocation of the content standards per competence.

Quality of allocating 72 content standards to eight competences was checked by analyzing whether the content standards per competence form a reliable scale (see Table 2 for results). Minimum criteria for scalability were item-rest correlations ≥ 0.35 and Cronbach's alpha ≥ 0.70. For all competences the accompanying content standards except one (item-rest correlation = 0.34) met these criteria of scalability. So, allocation of content standards to a preliminary rubric framework of performance standards has been successful.

3.2 Pilot of rubric framework

Thirdly, a pilot study was done. Participants were two researchers, two teacher educators, two team managers and two workplace coaches, all with experience in teaching nursing students reflecting skills in a traditional as well as in a competence-based nursing curriculum. They judged the rubrics to be clear, representative of the content standards and feasible in practice. They preferred a rubric framework with four instead of three performance levels because it makes more differentiated assessments possible.

As including four performance levels would result in a framework covering a broader continuum for teacher development, rubrics with four performance levels were developed. Moreover, a general description was added to each of the eight competences. See Table 3 for an example of a rubric for one of the competences. These adjustments resulted in the first version of the rubric framework which was input for the empirical performance standard-setting study, which will be described below.

Table 3. Rubric of competence 3 (first version)

Give information and instruction how to reflect General description: The teacher distinguishes between the objects and thinking activities of reflection which can be applied. Objects of reflection are: theory, practice and person (self). Thinking activities are: describing, analyzing, structuring, explaining, evaluating, concluding, attributing, formulating intentions. She* supports students by discussing, explaining, illustrating and demonstrating objects and thinking activities of reflection. She stimulates discussions between students and concludes together with them. In addition she pays attention to similarities and differences between theory and practice and the interplay between knowledge, skills and beliefs of students and the learning environment. She teaches students to apply thinking activities. She gives students reflection tasks and suitable material and asks questions to stimulate the use of thinking activities. She teaches students to ask each other and themselves such questions.				
	Rubric descriptions on four performance levels			
Levels → Rubric attributes ↓	Beginning	Developing	Proficient	Competent
Objects of reflection	Does not distinguish objects	Distinguishes theory, practice, self	Teaches students to distinguish objects	Lets students distinguish objects
Thinking activities	Does not distinguish thinking activities	Distinguishes (eight) thinking activities	Teaches students to distinguish thinking activities	Lets students apply thinking activities
Examples, demonstration how to reflect	Does not explain and does not give examples	Explains and gives examples	Demonstrates, stimulates discussion	Lets students lead discussions, concludes
Questions to stimulate reflection	Does not ask questions	Asks stimulating questions	Teaches students to ask each other questions	Teaches students to ask themselves questions
Reflection tasks	Does not give tasks	Gives stimulating tasks	Elucidates tasks	Suggests study tasks and material

*Where 'she' is used also 'he' can be read.

3.3 The study

3.3.1 Participants

Six nursing teams, from five schools and one workplace, voluntary participated in the rubric procedure (activity 4). These teams cooperate with the Dutch Foundation for Innovation of Vocational Education, which develops methods, materials and instruments to support students' growth in competence [25]. The participants were 33 teachers, three workplace coaches and four team managers (10 men and 30 women) who ranged in age from 26 to 63 years. The teachers had on average 11 years of work experience in nursing education.

3.3.2 Design and procedure

Per team three meetings were conducted between March 2008 and March 2009. These meetings were audio taped to determine procedural and internal validity evidence of the rubric procedure (activity 10). Procedural validity evidence concerned whether participants understood the rubric procedure and perceived the procedure as attainable in time, and whether they set performance standards. Internal validity evidence concerned consensus between participants on the content of the rubric framework and on the judgmental model.

In the first meeting a team was informed about the development process of rubrics. Also the goal and method of the standard-setting procedure were explained (activity 5). At the end of the first meeting teachers decided about their voluntary participation in the study. Next, each participant received the first version of the rubric framework by email.

In the second team meeting the rubric framework was explained and performance levels descriptions were illustrated. Participants could freely ask questions about the framework (activity 5). Then, to get acquainted with the rubrics, they used the performance descriptions to self-assess their competences. Next, participants discussed the performance descriptions and labels for performance levels in view of their relevance and acceptability. Further, suggestions and revisions deemed necessary given by former teams, were brought in if they were not mentioned by the present team. The purpose was to obtain clear and acceptable rubrics and labels on which consensus could be reached within and between teams (activity 9). This approach contributes to the internal validity (activity 10). Finally, participants determined a cut-off score per attribute independently from other participants by scoring the performance level they would consider as a "just sufficient" performance level for a starting teacher (activity 6). After this meeting mean cut-off scores were computed (activity 7).

The workplace team was unable to participate in the third meeting due to workload in their organization. Therefore the third meeting took place in only five teams (instead of six). Each team was informed about suggestions and revisions which had contributed to the second version of the rubrics. Then the team was asked whether they could agree with this second version. Further, each team was informed about their mean cut-off scores per rubric attribute. Moreover, these scores were compared with mean cut-off scores of all teams together. In addition, the team was informed about judgmental models they could apply and their essential differences (activity 4). After that, each team discussed which judgmental model the team members assessed as most suitable (activity 8). This discussion lasted until the team reached consensus. Finally, consensus on preferred cut-off scores and the judgmental model was member-checked by asking whether each participant supported the choices made (activity 9).

3.3.3 Data analysis

All audio tapes were used to analyze whether participants indicated that the rubric procedure was clear and the time schedule, three meetings of 90 minutes within one year, was attainable for each team. Moreover, a check was done whether each participant determined a cut-off score per attribute on paper. The audio tapes also were used to check suggestions and revisions made by the teams about the rubric descriptions and the rubric labels and whether consensus was reached (activity 10).

Participants' judgments of cut-off scores were analyzed for the degree of consensus on the performance standards. First, per rubric attribute means and standard deviations were computed of the cut-off scores of all participants. Then the degree of consensus was determined between participants on minimal acceptable performance levels (activity 7).

In addition, a check was done whether significant differences exist between teams in their judgments on cut-off scores before starting discussion about cut-off scores within teams. Therefore per rubric attribute mean cut-off scores and standard deviations were computed of all participants together and of six teams. Univariate analyses of variance were used to determine whether these teams differed significantly in their mean cut-off scores and $p = 0.05$ was considered to indicate a statistically significant difference. Consistency between and within teams contributes to internal validity evidence for the performance standards (activity 10).

Finally, reliability analyses were conducted to determine whether participants' cut-off scores per rubric attribute formed a scale per competence. Consistency between rubric attributes per competence was examined for all participants together by computing Cronbach's alphas. Minimum criteria for scalability were item-rest correlations ≥ 0.35 and Cronbach's alpha ≥ 0.70. Consistency between rubric attributes also contributes to internal validity evidence (activity 10).

4 Results

4.1 Procedural and internal validation

Procedural validity evidence was obtained because participants had no questions about or comments on the rubric procedure and the procedure was attainable within the three meetings arranged within a school year. Moreover, each participant judged the performance standards and determined cut-off scores.

The rubric framework was discussed in the second and third round of team meetings. In the second round the first version of rubric descriptions appeared to be fully clear to the six teams. They agreed to a high extent with proposed descriptions of performance levels of the rubrics. Teams had no comments on descriptions of the competences 'coaching' (competence 4) and 'feedback' (competence 5). For the other six competences they suggested minor changes. All suggestions were used to develop the second version of the rubric framework.

Two of the four labels of the performance levels generated some confusion. All teams thought that the performance label 'Developing' for the second level could not be used as a level because in vocational education the focus on life-long learning implies that development is a permanent process. The teams chose 'Basic' as new label. In addition, in every team some participants perceived the label 'Competent' of the fourth performance level as incorrect because competent cannot be the highest possible performance level. After a discussion about possible alternatives, agreement was reached in all teams about 'Expert' as new label. Each of six teams decided unanimously that the labels for the first and third performance levels needed no changes.

In the third round of meetings each participant agreed with the second version of the framework. Teams concluded that this version contained unambiguously formulated rubrics which are relevant and acceptable for teachers who support nursing students' reflection skills. Internal validity evidence was obtained because consensus was reached between and within teams.

4.2 Cut-off scores

Scoring performance levels of rubrics appeared to be feasible for participants. They individually scored for all 30 rubric attributes which performance level they would consider as "just sufficient". For all 30 rubric attributes most of the individual cut-off scores were on performance level 2. In four teams mean cut-off scores were ≤ 2.5 for most attributes (87% till 93%). In the fifth team this was 50% and the sixth (workplace) team 33%, due to level 4 scores of one participant in the team.

Mean cut-off scores of all participants together on all rubric attributes varied between M = 2.1 to M = 2.6. The lowest mean score (M = 2.1) was obtained on four attributes within three competences about subject matter and didactics: "distinguish and discuss reflection objects" (competence 3: instruction), "discuss patterns" (competence 5: feedback), and "support broad reflection" respectively "support deep reflection" (competence 6: assessment). The highest mean score (M = 2.6) was obtained on rubric attributes about applying verbal respectively non-verbal skills (competence 1: effective communication). For an overview of results, see Table 4.

Table 4. Reliability, means and standard deviations of rubric framework

Competence scales		**Cut-off scores**		
Competences	α	Rubric attributes for teaching reflection skills	M	SD
1 Effective communication with individual and group	0.84	1 apply verbal skills	2.6	0.6
		2 apply nonverbal skills	2.6	0.7
		3 discuss students' functioning	2.4	0.6
		4 evaluate students' functioning	2.3	0.5
2 Care for a positive learning climate	0.70	5 build a pedagogical relation	2.4	0.5
		6 attune the target group	2.5	0.5
		7 coach empathically	2.5	0.8
3 Give information and instruction how to reflect	0.86	8 distinguish and discuss reflection objects	2.1	0.5
		9 distinguish and apply thinking activities	2.2	0.7
		10 explain examples, demonstrate how to reflect	2.3	0.5
		11 ask questions to stimulate reflection	2.4	0.7
		12 give reflection tasks	2.2	0.7
4 Coach the learning process of how to reflect	0.88	13 give guidance	2.4	0.7
		14 use methodology	2.4	0.7
		15 teach how to develop an attitude	2.5	0.6
		16 coach students	2.3	0.6
5 Give feedback, teach how to receive and ask for feedback	0.90	17 give feedback	2.3	0.7
		18 receive feedback	2.4	0.8
		19 ask feedback	2.4	0.6
		20 discuss patterns	2.1	0.5
		21 make relations	2.2	0.5
6 Use criteria to assess reflections	0.86	22 support broad reflection	2.1	0.5
		23 support deep reflection	2.1	0.6
		24 support systematical reflection	2.5	0.7
		25 discuss constraints	2.2	0.6
7 Design and organize learning environment	0.94	26 care for materials	2.3	0.6
		27 work systematically	2.3	0.6
8 Work individually and in cooperation with colleagues	0.87	28 cooperate with colleagues	2.3	0.5
		29 professionalize	2.4	0.6
		30 be a role model	2.3	0.6

4.3 Choice of a judgmental model

Five teams discussed mean cut-off scores of their team and compared them with mean cut-off scores of other teams. For most of the rubric attributes mean cut-off scores per team were between M = 2.0 and M = 2.7. On three of 30 attributes

mean cut-off scores differed significantly between teams. These concerned attribute "apply verbal skills" of competence 1: $F(5) = 3.2$, $p = 0.02$, attribute "distinguish thinking activities" of competence 3, $F(5) = 3.6$, $p = 0.01$, and attribute "discuss patterns" of competence 5, $F(5) = 2.8$, $p = 0.03$. Effect sizes, calculated using eta squared, were respectively 0.05, 0.03 and 0.09. Post-hoc comparison using the Bonferroni test indicated that mean scores only were significantly different on attribute "distinguish thinking activities" ($p = 0.047$, $p = 0.030$, $p = 0.010$).

Five teams also discussed rubric attributes examples in which compensating within the same competence might be possible. However, no rubric attributes could be found for which teams thought performance level 1 was enough. Discussions per team led to consensus on level 2 as the "just sufficient" performance level for each attribute within each of the eight competences. So these teams decided to apply a conjunctive judgmental model for each competence.

4.4 Scale reliabilities

Using individual scores of participants, scale analyses showed that for all eight competences all rubric attributes per competence had item-rest correlations ≥0.35 and all Cronbach's alphas were ≥ 0.70, most of them falling between 0.84 and 0.88. The lowest alpha was 0.70 (competence 2) and the highest 0.94 (competence 7).

5 Discussion

In the study the choice of rubrics (activity 1) was effective because it stimulated discussions about relevance and acceptability of teachers' competences and rubric attributes. The developed rubric framework of performance descriptions (activity 3) appeared to be relevant and acceptable for all six nursing institutes. Participants formed a panel of representative judges of senior secondary nursing education (activity 4). These judges received information about rubrics, the standard-setting procedure and judgmental models and practiced setting standards (activity 5).

Participants gave, independently of each other, cut-off scores on the 30 rubric attributes (activity 6) [14, 18, 19]. Mean scores per team were mostly ≤ 2.5 (activity 7). More extreme scores of four participants only led to a significant result on one attribute within competence 3. These findings show a consistency between the cut-off scores of the teams. Consistency in judgments contributes to internal validity evidence (activity 10) [14, 19].

Discussions were held until consensus was reached on the judgmental model (activity 8). Teams decided for all eight competences that the minimum performance level would be level 2 for each rubric attribute. So teams chose a conjunctive model for each competence [14, 18]. Moreover, between all teams full agreement was obtained.

During the standard-setting procedure activities were evaluated and documented (activities 9 and 10). Our procedure for setting performance standards in ten activities appeared to be useful, according to the participants and led to the intended outcomes: performance standards and a judgmental model convenient for assessing teachers' competences to support nursing students' development of reflection skills. These findings contribute to procedural validity evidence (activity 10).

A limitation in the study was that meetings for the six teams were held subsequently. Changes deemed necessary by a team were discussed in the other teams, to take their bearings into account and come to performance levels that are appropriate, unambiguously formulated and clearly distinguishable. Although we gave all teams the same information in the same sequence, the specific sequence used might have influenced the comments of later teams.

6 Conclusion

Procedural and internal evidence both justify the conclusion that the rubric procedure, as used in this study, led to consensus within and between teacher teams in nursing education. Also, rubrics are convenient for developing and validating performance standards. This study generates practical implications, since teacher teams perceived the rubrics as

relevant and acceptable for supporting nursing students' reflection skills development. So these rubrics can be used in teacher training on giving students feedback, to support nursing students' reflection skills.

Further research can be done to explore whether the rubrics can be employed to prepare and formatively assess reflection conversations, and how teachers develop their competences to support students' reflection skills development. Research of the use of rubrics in nursing education may give insights into whether analytical scoring is suitable to teachers. Moreover, research can focus on the generalization of rubrics to other domains e.g. social work and teacher education. Future studies might also provide external validity evidence for the rubrics.

Sources of support

This research is funded by NWO-PROO (Project 411-06-311). The project is part of a Dutch national cooperation of Eindhoven University, Leiden University, and Utrecht University.

References

[1] Biesma RG, Pavlova M, Vaatstra R, Van Merode GG, Czabanowska K, Smith T. Generic Versus Specific Competencies of Entry -Level Public Health Graduates: Employers' Perceptions in Poland, the UK, and the Netherlands. Advances in Health Sciences Education. 2008; 13: 325-343. PMid:17151830 http://dx.doi.org/10.1007/s10459-006-9044-0

[2] Harden RM, Crosby J. AMEE Guide No 20: The good teacher is more than a lecturer- the twelve roles of the teacher. Medical Teacher. 2000; 22: 334-347. http://dx.doi.org/10.1080/014215900409429

[3] Schön DA. The reflective practitioner. How professionals think in action. Aldershot Hants: The Academic Publishing Group. 1991.

[4] Govaerts MJB. Educational competencies or education for professional competence? Medical education. 2008; 42: 234-236. PMid:18275410 http://dx.doi.org/10.1111/j.1365-2923.2007.03001.x

[5] Mann K, Gordon J, MacLeod A. Reflection and reflective practice in health professions education: a systematic review.Advances in Health Sciences Education. 2009; 14: 595-621. PMid:18034364 http://dx.doi.org/10.1007/s10459-007-9090-2

[6] Korthagen. F, Wubbels T. Evaluative Research on the Realistic Approach and on the Promotion of Reflection. In Linking practice and theory: the pedagogy of realistic teacher education. F Korthagen, J Kessels, B Koster, B Lagerwerf and T Wubbels, eds. Mahwah, NJ: Lawrence Erlbaum Associates. 2001; 88-107.

[7] O Donovan M. Implementing reflection: Insights from pre-registration mental health students. Nurse Education Today. 2007; 27: 610-616. PMid:17081658 http://dx.doi.org/10.1016/j.nedt.2006.09.001

[8] Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. Academic Medicine. 2004; 79: 70-81. PMid:15383395 http://dx.doi.org/10.1097/00001888-200410001-00022

[9] Shute VJ. Focus on formative feedback. Review of Educational Research. 2008; 78: 153-189. http://dx.doi.org/10.3102/0034654307313795

[10] Sadler DR. Formative assessment and the design of instructional systems. Instructional Science. 1989; 18: 119-144. http://dx.doi.org/10.1007/BF00117714

[11] Sadler DR. Beyond feedback: Developing student capability in complex appraisal. Assessment and Evaluation in Higher Education. 2010; 35: 535-550. http://dx.doi.org/10.1080/02602930903541015

[12] European Commission Education & Training, Teacher Education [Internet]. 2011. Available from: http://ec.europa.eu/education/school-education/doc832_en.htm

[13] National Board for Professional Teaching Standards [Internet]. 2011. The Standards. Available from: http://www.nbpts.org/the_standards

[14] Hambleton RK, Pitoniak MJ. Setting Performance Standards. In Educational Measurement (4th edition). RL. Brennan, ed. Westport: Praeger. 2006; 433-469.

[15] Boursicot KA, Roberts T, Pell G. Standard Setting for Clinical Competence at Graduation form Medical School: A Comparison of Passing Scores across Five Medical Schools. Advances in Health Sciences Education. 2006; 11: 173-183. PMid:16729244 http://dx.doi.org/10.1007/s10459-005-5291-8

[16] Norcini JJ. The metric of medical education. Setting standards on educational tests. Medical Education. 2003; 37: 464-469. PMid:12709190 http://dx.doi.org/10.1046/j.1365-2923.2003.01495.x

[17] Harrison R, Mitchell L. Using outcomes-based methodology for the education, training and assessment of competence of healthc are professionals. Medical Teacher. 2006; 28: 165-170. PMid:16707298 http://dx.doi.org/10.1080/01421590500271308

[18] Berk RA. Standard Setting: The Next Generation (Where Few Psychometricians Have Gone Before!). Applied Measurement in Education 1996; 9: 215-235. http://dx.doi.org/10.1207/s15324818ame0903_2

[19] Kane M. Validating the Performance Standards Associated With Passing Scores. Review of Educational Research. 1994; 64: 425-461.

[20] Arter J, Chappuis J. Creating & recognizing Quality rubrics. New Jersey: Pearson Education Inc. 2007.

[21] Hafner JC, Hafner PM. Quantitative analysis of the rubric as an assessment tool: An empirical study of student peer-group rating. International Journal of Science Education. 2003; 25: 1509-1528. http://dx.doi.org/10.1080/0950069022000038268

[22] Lane S, Stone CA. Performance Assessment. In Educational Measurement (4th edition). RL. Brennan, ed. Westport: Praeger. 2006; 387-431.

[23] Ministry of Education, Culture and Science [Internet]. 2007. Wet BIO. [Law on educational professions]. Available from: http://www.rijksoverheid.nl/documenten-en-publicaties/vragen-en-antwoorden/hoe-wordt-de-bekwaamheid-van-docenten-bewaa kt.html (18 October 2011 date last accessed).

[24] Dekker-Groen AM, Van der Schaaf MF, Stokking KM. Teacher Competences required for developing Reflection Skills of nursing students. Journal of Advanced Nursing. 2011; 67: 1568-1579. PMid:21332576 http://dx.doi.org/10.1111/j.1365-2648.2010.05591.x

[25] Klatter EB. Evalueren, Beoordelen en Kwalificeren van competentie ontwikkeling. Visie document. [Evaluating, assessing and qualifying competence development. Vision document] Nijkerk, The Netherlands: Stichting Consortium Beroepsonderwijs groep Zorg & Welzijn. 2007.