Chapter 3. Measuring personal reflection: The Groningen Reflection Ability Scale (GRAS)

De zekerheid van de natuur is de som van vele onzekerheden, die allemaal te meten zijn. Hoe meer onzekerheid je meet, des te zekerder ben je van de uitkomst. – Gerrit Krol

Abstract

Aim: Personal reflection is important for acquiring, maintaining and enhancing balanced medical professionalism. A new scale, the Groningen Reflection Ability Scale (GRAS), was developed to measure the personal reflection ability of medical students.

Method: Explorative literature study was conducted to gather an initial pool of items. Item selection took place using qualitative and quantitative methods. Medical teachers screened the initial item-pool on relevance, expert-analysis was used for screening the fidelity to the criterion and large samples of medical students and medical teachers were used to investigate the psychometric characteristics of the items. Finally, explorative factor analysis was used to investigate the structure of the scale.

Results: The psychometric quality and content validity of the GRAS are satisfactory. The items cover three aspects of personal reflection: self-reflection, empathetic reflection and reflective communication. The 23-item scale proved to be easy to complete and to administer.

Conclusion: The GRAS is a practical measurement instrument that yields reliable data that contribute to valid inferences about the personal reflection ability of medical students and doctors, both at individual and group level.

Introduction

Doctors have to solve complex problems and they are expected to do that in cooperation with patients and their families, in multi-professional collaboration, in a professional and personal way, and while learning from their experiences. This complexity, inherent to clinical practice, demands not only skillful behaviour and sophisticated types of clinical reasoning and scientific reflection, but also personal
reflection. The aim of this study was to construct a practical scale for measuring the ability of personal reflection. This article describes the development of the Groningen Reflection Ability Scale (GRAS), the psychometric and substantial characteristics of the items, and the structure of the scale.

**Box 1. Practice points**

- Personal reflection is an important construct in medicine that can be measured.
- Whereas available instruments are mainly focused on critical thinking regarding well-defined clinical problems, the GRAS is measuring personal reflection regarding important multi-facetted problems in medicine.
- The GRAS is a one-dimensional scale, covering three relevant aspects of personal reflection: Self-reflection, empathetic reflection and reflective communication.
- The GRAS yields reliable data that contribute to valid inferences about the personal reflection ability of medical students and doctors, both at individual and group level.

Reflection is required for adapting professional functioning to the patients’ needs or to new circumstances when there are no obvious solutions, for transformation of new knowledge and into practice, or for life-long personal and professional learning. In literature large numbers of forms and functions of reflection are distinguished and promoted. For example, mindful practice (Epstein 1999), coping with ill-structured problems (King & Kitchener 1994), awareness of emotions (Boud & Walker 1993), assumptions (Van Manen 1990), or morality (Kohlberg 1984), learning from experience (Schön 1987), deep learning (Marton et al. 1984) or critical learning (Mezirow 1992), and the assessment of integral learning outcomes (Shumway & Harden 2003). Reflection, especially personal reflection on experience, is seen as a key factor for acquiring and maintaining balanced professionalism along the continuum of medical education (General Medical Council 1999; Irvine 1999; Simpson et al. 2002; BIM, ACP-ASIM & EFIM 2002; CanMEDS 2000/2005; Crues & Crues 2006). However, in contrast to the richness of opinions and efforts for implementing reflection in medical education, the attention given to instruments to measure reflection is relatively poor (Kember & Leung 2000; Arnold 2002). The available instruments are mainly focused on clinical reasoning and critical thinking regarding well-defined problems (King & Kitchener 1994), or on specific aspects of reflection in medical and nursing education, such as identity formation (Niemi 1997), moral reasoning (Kohlberg 1984), or reflective writing (Wong et al. 1995).
In order to structure this confusing amount of types of reflection and to clarify the focus of this study, we distinguish three cognitive-emotional levels of the doctor as a reflective practitioner. First of all there is clinical reasoning, the problem- and patient-oriented understanding, judgment and decision, with the key function of problem solving. Then comes scientific reflection as the predominant type of critical thinking in medicine: The critical appraisal of literature and own practice, preferentially grounded in evidence and clinical epidemiology (Sackett et al., 1991), with the key function of optimizing scientifically-based clinical decisions or evidence based medicine (EBM). Finally there is personal reflection on experience, with the key function of optimizing balanced professionalism in medicine.

According to Donald Schön’s theory (1987) of the reflective practitioner, clinical reasoning can be conceptualized as ‘reflection-in-action’, and scientific thinking as the prototype of ‘reflection-on-action’ in medicine. We have a preference to describe personal reflection as ‘reflection-on-experience’ to emphasize the direction of reflective attention to the process of sense-making in medical practice, and to the dynamics of rational and irrational thoughts and emotions, assumptions and beliefs in that process.

Personal reflection in medicine can be described as the exploration and appraisal of one’s own and other’s experiences, thus clarifying and creating meaning, for the benefit of balanced functioning, learning and development. This definition of personal reflection integrates some key aspects complementary to scientific reflection: Mindfulness rather than intellectualism; attention for experience rather than action; appropriate communication, such as handling a dialogue and feedback rather than debate and discussion; clarifying the process of sense-making rather than problem solving; resulting in transforming or confirming one’s own perspective on professional practice and identity. Such a perspective fits in a multi-levelled model of balanced reflective practice, with the mentioned cognitive-emotional levels of clinical reasoning (doctor as expert), scientific reflection (doctor as scholar) and personal reflection (doctor as person).

It cannot be denied, however, that in the pragmatic problem-solving culture of medical practice and education personal reflection often was, and is, seen as a hard to conceive subjective and abstract issue. Personal reflection is usually seen more as a private attitude or personality trait of doctors, and thus by definition positioned outside the medical toolkit, and the professional and educational domain (Coulehan 2005). One explication for this separation is the fact that scientific reflection, obviously of vital importance for the medical profession and strongly established in medical education (Freidson 1994), is also so predominant that it often seems to be prototypical for reflection in medicine. This would be a mistake and a threat for
balanced medical professionalism by constituting an overall one-dimensional technical view, regarding relatively well-structured clinical problems. And thus disregarding important less structured problems in patient care, inter-collegial cooperation and own functioning. For balanced medical professionalism it is of vital importance that the doctor keeps an awareness of the tacit constitutive processes of attention, selection and socialization (Epstein, 1999). We agree with Hilton and Slotnick (2005), who state that ‘skills, knowledge and experience are necessary for professionalism, but sophisticated reflection on the doctor’s part is also required to produce insights enabling the individual to better address the needs of patients specifically, and society generally’. We use the term personal reflection for this mode of reflection. Taking personal reflection seriously, as a real mental hygienic component of balanced medical functioning, that can be practiced and developed within certain personal and cultural boundaries, implies the need for instruments by which it can be measured.

Methods

Participants and methods of the item selection procedure
First an initial item pool was constructed with descriptions of reflection gathered from literature, personality theory and educational practice. Then, item selection took place using qualitative and quantitative methods: Medical teachers screened the initial item pool on fidelity to the criterion, expert-analysis was used for screening the relevance, large samples of medical students and medical teachers were used to investigate the psychometric characteristics of the items, and explorative factor analysis was used to investigate the structure of the scale. This item selection procedure will be described now step by step (see Table 1).

Start. The principal researcher started with the screening of descriptions of reflection from different sources. Descriptions of reflective ability, behaviour and personal characteristics as mentioned in the literature on reflection, in lists used in the Five Factor personality theory (De Raad & Doddema 1999; Hendriks 1997), and from own educational practice were accumulated and screened with the criterion: ‘Which descriptions are associated with personal reflection in a medical context?’ This resulted in a pool with items that could be scored on a 5-point Likert scale, with the question: To which extent does this description applies to you? (1 meaning ‘totally disagree’ until 5 ‘totally agree’).

Step 1. The resulting list of items was discussed and screened at face value by 11 colleagues participating in curriculum development and medical teaching. Some items were reformulated based on experiences in medical education and in a joint
multi-centre project on reflective professional behaviour. Then the list was judged by 20 faculty members (both medical doctors and social scientists) who were participating in a curriculum project aiming at developing competent reflective doctors. Ten respondents were asked to judge the items as ‘characteristic for the good doctor’, and 10 respondents as ‘characteristic for the bad doctor’. Non-discriminating items were removed.

*Step 2.* The resulting list of items was completed by a sample of 350 first-year medical students of Groningen University and 38 medical teachers (general practitioners, participating as facilitator). The facilitators asked the students to complete the scales during one of their coach-group sessions. This measurement was followed by psychometric analysis. For item-variance a criterion standard deviation of 0.75 was taken, whereas a SD of 1.0 on a 5-point Likert scale is usually seen as sufficient (Nunnaly 1967). Items with a SD< 0.75 were removed. The item intercorrelation was controlled by checking items that systematically scored low or negative. The item discrimination was calculated with using item-rest correlations.

<table>
<thead>
<tr>
<th>Steps and Goal</th>
<th>Scale N items</th>
<th>Participants</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start</td>
<td>3456</td>
<td>1 Principal researcher</td>
<td>Using literature</td>
</tr>
<tr>
<td>Step 1. Face validity analysis</td>
<td>81</td>
<td>11 Researchers and colleagues 20 Faculty members</td>
<td>Using experience</td>
</tr>
<tr>
<td>Step 2. Psychometric item analysis</td>
<td>61</td>
<td>350 1st year students 38 Medical teachers</td>
<td>Measurement</td>
</tr>
<tr>
<td>Step 3. Face validity analysis</td>
<td>50</td>
<td>38 Medical teachers (same group) 21 Experts</td>
<td>Measurement Expert-analysis</td>
</tr>
<tr>
<td>Step 4. Psychometric structure analysis (construct validation)</td>
<td>23</td>
<td>583 1st ~6th years students 38 Medical teachers (same group) 14 Medical skills teachers 102 1st ~6th years students 9 (next study year)</td>
<td>Measurement</td>
</tr>
</tbody>
</table>

*Step 3.* The resulting item list was completed again by the same 38 medical teachers. The results were used for both psychometric and expert analysis. After psychometric analysis, the items were analysed twice by experts. Firstly 19
experienced medical staff members independently judged the items, with the criterion: ‘How relevant is this item for the measurement of reflection?’, on a 5-point Likert scale (1 meaning ‘not relevant at all’ until 5 ‘very relevant’). Items with a mean score >4 and a SD < 1 on these ratings were interpreted as highly relevant and were kept in the list. Three well-known experts from outside the faculty then analysed the reduced list. They scored each item for relevance on paper, followed by a semi-structured focus-group meeting, in which each item was discussed critically for relevance, conceptual clarity and social desirability. In some cases of doubt psychometrically less satisfactory items with a high score at relevance were reformulated and kept in the list, because relevance was considered to be of prime importance.

Step 4. For psychometric item and structure analysis two measurements were completed. The first measurement, with a sample of 583 first-years to sixth-year medical students and 38 medical teachers, the same as in step 1; and 14 experienced teachers in medical skills, was used for psychometric item analysis. This was needed because of reformulations of some items after the expert-consultation in step 3. The second measurement, with a sample of 1029 first-year to sixth-year medical students in the next study year, was used for psychometric structure analysis. The scale was constructed as a one-dimensional scale, but we were interested in exploring possible subsets of items, that are distinct from other item groups and homogeneous within a group, as pointers of the content validity of personal reflection. Therefore explorative factor analysis (with varimax rotation) was used.

Results

Results of the item selection procedure
The principal researcher started the item collection procedure by screening about 100 descriptions from literature on reflection and 3456 personal characteristics from lists used in the Five Factor personality theory, resulting in a pool of 81 items. After screening fidelity to the criterion by researchers and faculty members (step 1) 20 items were removed. The measurement with the resulting 61-item list (step 2) resulted in the following psychometric qualifications. Item difficulty: The average of the items was between 3 and 4. Item-variance: With a criterion standard deviation of 0.75 33 items were removed. The item-discrimination: 1 item with an item-rest correlation of \( r_{i-r} < 0.20 \) was discarded. This psychometric analysis resulted in 27 psychometrically satisfying and 34 less satisfying items. However, 23 of the 34 psychometrically less satisfying items were judged as relevant by the
researchers from a substantial and educational point of view. Consequently these 23 items were reformulated and kept in the list, resulting in a 50-item list.

The psychometric analysis (step 3) resulted in 27 psychometrical satisfying items. During the fidelity to the criterion analysis by faculty experts 4 of these 27 psychometrical satisfying items were judged as weak on relevance and removed, as a consequence of the prime importance of relevance. 6 of the psychometrically unsatisfying items were kept because they were judged as relevant, resulting in a 29-item list. The fidelity to the criterion analyses by three extern experts of this list resulted in a 23-item list. This 23-item list was used for two final measurements (step 4) for psychometric item and structure analysis. This blended procedure of psychometric and expert analyses, reductions and reformulations, resulted in the definitive Groningen Reflection Ability Scale (GRAS), containing 23 items – of which 14 items remained from the initial list of 61 items (step 1), and 19 items were new or re-formulated.

**Characteristics of the definitive scale**
The GRAS has a satisfactory internal consistency: A Cronbach’s alpha of 0.83 for the 1st measurement (step 4) and 0.74 for the 2nd measurement (step 4), an item-difficulty (between 4.23 and 3.50) and item variance (between 1.04 and 0.67) (all related to the Dutch items). The scale proved to be easy to administer and to complete, within 10 minutes. The scores can be calculated without time-consuming coding procedures.

The initial content validation, by means of structure analysis using explorative factor analysis, was based on three criteria: Eigenvalues, Scree test and substantial criterion. (a) Using an eigenvalue criterion of >1 resulted in 7 factors that were difficult to interpret. (b) However, with the Scree test, the first factor could be isolated with a jump of 10.5% explained variance between factor 1 and 2. In addition, the first three factors could be isolated with a smaller jump of 0.8% explained variance between factor 3 and 4. (c) Using the substantial criterion, the first three factors could be interpreted as three relevant groups of items (Table 2): Self-reflection (10 items): Introspection, exploration, understanding and appraisal of experiences; empathetic reflection (6 items): Replacement in and taking into consideration the situation of others, openness to different ways of thinking, contextual understanding and appraisal; and reflective communication (7 items): Reflective behaviour, openness for feedback and discussion, taking responsibility for own statements and actions, ethical accountability.

**Table 2.** Summary of principal components (with eigenvalues) contributing to each subgroup of items as a result of explorative factor analysis of the GRAS structure.
### Table 1: Factor Loadings and Communalities

<table>
<thead>
<tr>
<th>Factors</th>
<th>I</th>
<th>II</th>
<th>III</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(3.98)</td>
<td>(1.56)</td>
<td>(1.47)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Item</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>I take a closer look at my own habits of thinking</td>
<td>.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I want to know why I do what I do</td>
<td>.59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I find it important to know what certain rules and guidelines are based on</td>
<td>.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I want to understand myself</td>
<td>.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I am aware of the emotions that influence my thinking</td>
<td>.51</td>
<td>.35</td>
<td>.31</td>
</tr>
<tr>
<td>I am aware of the emotions that influence my behaviour</td>
<td>.44</td>
<td>.33</td>
<td>.38</td>
</tr>
<tr>
<td>I am able to view my own behaviour from a distance</td>
<td>.40</td>
<td></td>
<td>.25</td>
</tr>
<tr>
<td>I test my own judgments against those of others</td>
<td>.39</td>
<td></td>
<td>.18</td>
</tr>
<tr>
<td>I can see an experience from different standpoints</td>
<td>.36</td>
<td></td>
<td>.28</td>
</tr>
<tr>
<td>I am aware of the cultural influences on my opinions</td>
<td>.36</td>
<td></td>
<td>.19</td>
</tr>
<tr>
<td>I am aware of the possible emotional impact of information on others</td>
<td>.56</td>
<td></td>
<td>.42</td>
</tr>
<tr>
<td>I can empathize with someone else’s situation</td>
<td>.56</td>
<td></td>
<td>.22</td>
</tr>
<tr>
<td>I am aware of my own limitations</td>
<td>.46</td>
<td></td>
<td>.22</td>
</tr>
<tr>
<td>I reject different ways of thinking</td>
<td>.45</td>
<td></td>
<td>.30</td>
</tr>
<tr>
<td>Sometimes others say that I do overestimate myself</td>
<td>.45</td>
<td></td>
<td>.37</td>
</tr>
<tr>
<td>I am able to understand people with a different cultural / religious background</td>
<td>.42</td>
<td></td>
<td>.25</td>
</tr>
<tr>
<td>I do not like to have my standpoints discussed</td>
<td>.57</td>
<td></td>
<td>.33</td>
</tr>
<tr>
<td>I sometimes find myself having difficulty in illustrating an ethical standpoint</td>
<td>.51</td>
<td></td>
<td>.30</td>
</tr>
<tr>
<td>I am accountable for what I say</td>
<td>.50</td>
<td></td>
<td>.34</td>
</tr>
<tr>
<td>I take responsibility for what I say</td>
<td>.49</td>
<td></td>
<td>.35</td>
</tr>
<tr>
<td>I am open to discussion about my opinions</td>
<td>.33</td>
<td>.46</td>
<td>.33</td>
</tr>
<tr>
<td>I sometimes find myself having difficulty in thinking of alternative solutions</td>
<td>.46</td>
<td></td>
<td>.21</td>
</tr>
<tr>
<td>I do not welcome remarks about my personal functioning</td>
<td>.31</td>
<td>.44</td>
<td>.32</td>
</tr>
</tbody>
</table>

All factor loadings > 0.30 are reported. † after Varimax rotation. Factor 1: Self-reflection. Factor 2: Empathetic Reflection. Factor 3: Reflective Communication.

Table 2 shows the item loadings for all the three factors and the communalities (h²). The eigenvalues of each factor are: Factor 1, Self-reflection: 3.98; factor 2,
Chapter 3. The GRAS

Empathetic Reflection: 1.56; and factor 3, Reflective Communication: 1.47. The correlations between the three factors are: 0.62 for factor 1 ~ factor 2; 0.57 for factor 1 ~ factor 3; and 0.42 for factor 2 ~ factor 3. The Kaiser–Meyer–Olkin (KMO) statistic as ‘goodness of fit’ is 0.78 (p < 0.001). Measured by the KMO statistics, sampling adequacy predicts if data are likely to factor well. KMO overall should be 0.60 or higher to proceed with factor analysis. The total variance explained (30%), the eigenvalues and the factor loadings are an indication that the three factor model is not working well. As a consequence the communalities, measuring the percent of variance in a given variable explained by all the factors jointly, are low. The correlations with attenuation correction between each of the item groups are high enough to support the interpretation of one-dimensionality. These psychometric results indicate that the GRAS is a one-dimensional scale with three relevant aspects of that dimension.

Discussion

The aim of this study was to develop a scale to measure one mode of reflection in medicine, personal reflection. The study resulted in a 23-item scale of the GRAS (on a 5-point Likert scale) which proved to be easy to administer and to complete (within 10 minutes). Scores can be calculated without time consuming coding procedures.

According to the standards for educational and psychological testing (American Educational Research Association et al. 1999), the GRAS (with Cronbach’s alpha’s of 0.83 and 0.74) can be regarded as ‘good’ for less important decisions at an individual level (0.70 ~ 0.80) and a group level (0.60 ~ 0.70), and ‘not quite sufficient reliable’ for important decisions at an individual level (0.80 ~ 0.90). Whereas other scales only focus at one specific aspect (Niemi 1997; Wong et al. 1995) or at critical reflection on well-defined clinical problems (King & Kitchener 1994), the GRAS focuses at personal reflection on experience and less structured problems.

The GRAS is a one-dimensional scale. The conceptual argument is the fact that the GRAS was developed as a one dimensional scale and item selection took place in one domain of theory and literature about reflection. The three factors, being a result of explorative factor analysis, must therefore be interpreted primarily as facets of one dimension. The psychometric argument is that the proportion of variance represented by the first principal component is a good approximation of the proportion of variance represented by the first common factor (Falissard, 1999), which is supported by the correlations of each of the three factors. The factors are
Chapter 3. The GRAS

not independent enough to use as separate scales and thus as separate scores of individuals with practical consequences. This means that in practice a one-GRAS-score is leading. On the other hand, the three GRAS factors are probably distinctive facets of the concept of reflection in medicine.

The content validity of the scale is satisfactory, because the items are grounded in reflection literature, and covering three substantive aspects of personal reflection in the context of medical practice and education: Self-reflection, Empathetic Reflection; and Reflective Communication. We see self reflection as the introspective aspect of personal reflection: The careful exploration and appraisal of experience, as a prerequisite for framing or reframing one’s thoughts, feelings, beliefs, norms or methods; empathetic reflection as the social, inter-subjective extension of self-reflection: Contextual understanding and appraisal, i.e. empathetic placement in and thinking about the position of others, such as patients and colleagues; and reflective communication as the behavioural expression of both self-reflection and empathetic reflection, for example the handling of feedback or a dialogue, or dealing with interpersonal differences.

The GRAS can be used in medical education for programme evaluation: Effect measurements on the reflection ability of medical students and doctors. The GRAS measures not only the effect of one course, but moreover to which extent the curriculum influences the growth-curve of medical students as reflective practitioners over a prolonged time, both at individual and group level. The GRAS can also be used for cross-sectional comparison between groups of medical students from different curricula, cultures and language, or between doctors from different medical expert disciplines.

A possible limited and paradoxical aspect of the GRAS is its self-rated character. Applicants are asked to judge their own reflection ability, which presupposes already a certain degree of self-reflection and self-observation. Although there is research demonstrating that it is a hard task to self-assess ones performance adequately (Norman, 1999; Kruger & Dunning 1999:), self judgments on personal characteristics do not automatically appear less accurate than peer judgments (Hofstee et al. 1998).

Although the content validity of the GRAS is satisfactory, further research can explore the external validity by comparing the GRAS with familiar reflection scales and with reflective criterion behaviour.

Further research about the varied forms and functions of reflection in medicine is needed. In the development of the GRAS we had a restricted focus on personal
reflection, conceptualized as one dimension. The GRAS can be used in combination with other measures, in a multi-method application, in order to capture the richness of reflection in medicine, both for practical and theoretical purposes. In further research the possible extension of the scale in a three factor model can be explored. We agree with Falissard (1999) that a priori unidimensionality is not true because, in our case, the reality of reflection in medicine does not correspond to any rigorous mathematical model. Reflection in medicine is a multidimensional construct that reflects the complexity of cognitive-emotional and meta-cognitive processes in medicine.

Conclusion

The GRAS is a practical measurement instrument that yields reliable data that contribute to valid inferences about the personal reflection ability of medical students and doctors, both at individual and group level. The GRAS is a one-dimensional scale with three relevant aspects of that dimension: Self-reflection, empathetic reflection and reflective communication. The 23 items on a 5-point Likert scale are easy to complete, resulting in a one-GRAS score. Scores can be calculated without time-consuming coding procedures.

References


De Raad B, Doddema M. 1999. *Book of Traits: The Language of Stable Characteristics of People and their Behaviours* (Groningen, University of Groningen, the Netherlands).


General Medical Council 1999. The Doctor as Teacher (London, General Medical Council).