The Suitability of Global Rating Scales to Monitor Arthroscopic Training Progress

JJ Stunt*, GMMJ Kerkhoffs1, B van Ooij2, IN Sierevelt1, MU Schafroth1, CN van Dijk1, JDragoo3 and GJM Tuijthof1,2

1Department of Orthopedic Surgery, Academic Medical Centre, The Netherlands
2Department of BioMechanical Engineering, Delft University of Technology, The Netherlands
3Department of Orthopedic Surgery and Sports Medicine, Stanford University School of Medicine, USA

*Corresponding author: Jonah Stunt, Department of Orthopedic Surgery, Academic Medical Centre, Orthotrauma Research Center Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands, Tel: 0205662173/9117, E-mail: j.j.stunt@uva.amc.nl

Abstract

Purpose: As developing arthroscopic skills is challenging and training time for residents is limited, arthroscopic skill competency of residents should be measured. Assessment tools, such as Global Rating Scales (GRS), have been developed for structured, objective feedback and to assess learning curves. The goal of this study is to assess known-groups and convergent validity of these scales, to evaluate the suitability of these scales to monitor training progress of residents.

Methods: Knee arthroscopies and ACL reconstructions performed by residents were supervised and assessed, using both GRS questionnaires. The estimates of the parameters were used to study the relationship between year of residency and each GRS score, and between the number of previously performed arthroscopies and each GRS score. Pearson correlation coefficient between GRS scores were calculated to measure convergent validity. A Bland-Altman plot with a paired t-test was constructed to evaluate the agreement between GRS I and II.

Results: Mixed model analysis revealed a significant increase (p < 0.001) per year of residency on both GRSs (8.1 points (95% CI: 6.3-9.9) and 9.2 points (95% CI: 7.4-11.2) respectively). Significant increases per performed arthroscopy were also observed for both GRSs (p < 0.001) (0.14 (95% CI: 0.09-0.18) and 0.13 points (95% CI: 0.08-0.2) respectively). Scores for ACL reconstructions were significantly lower (p < 0.001) than for standard knee arthroscopies (12.5 and 13.0 points respectively, p < 0.001). The Pearson correlation coefficient between GRS I and GRS II scores was high (0.94). The Limit of Agreement was 11 points.

Conclusion: GRS I and GRS II demonstrate sufficient construct validity. However, they seem not sufficiently sensitive and consistent to establish individual learning curves. Both scales are suitable to objectively evaluate global progress of residents in the operating room when acquiring arthroscopic skills, in particular on group level.

Keywords

Global rating scales (GRS), Objective assessment, Learning curves, Arthroscopic training

Introduction

Since arthroscopic surgery has several advantages compared with open surgery, it has become the most performed procedure in orthopaedic surgery [1-6]. However, as developing arthroscopic skills is challenging [7-10] and training time for residents is limited [9,11-13], professional societies have requested arthroscopic skill competency of residents to be assessed to improve patient safety [4,9]. Assessment of skills by expert surgeons is sensitive to the subjective opinion of the assessor, which might compromise fair judgment [14].

To overcome this issue, the formulation of criteria and proficiency levels for evaluation of arthroscopic skills is recommended.

Assessment tools for monitoring technical skills in the operating theatre, such as Global Rating Scales (GRS), have been developed for structured, objective feedback and to assess learning curves [1]. Previous research investigated whether GRS are valid, reliable tools to objectify resident performance in surgery; feasibility, face validity, content validity, construct validity and reliability have been demonstrated for various Global Rating Scales [4,15-23]. Two Global Rating Scales that have been specifically proposed for feedback during arthroscopic training, are the Basic Arthroscopic Knee Skill Scoring System [7] and the Orthopaedic Competence Assessment Project [24]. Insel and co-workers developed a GRS to assess diagnostic knee arthroscopies and partial meniscectomies on cadaver knees (GRS I) [7]. This GRS has demonstrated validity, but only on cadavers, performing basic arthroscopic tasks [7]. Howells and co-workers combined the Orthopaedic Competence Assessment Project and the Objective Structured Assessment of Technical Skill (OSATS) (GRS II) to test arthroscopic simulator training on a bench-top knee simulator [24]. Validity of this GRS has however not yet been assessed.

Since both GRSs are not validated in a clinical context, the goal of this study is to assess validity of these scales during training of arthroscopic skills in the operating room on real-life patients. In the absence of a gold standard, both known-groups and convergent validity are investigated to assess the suitability of these scales to monitor training progress of residents.


Received: August 16, 2015; Accepted: May 14, 2016; Published: May 17, 2016

Copyright: © 2016 Stunt JJ, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
### Materials and Methods

#### Participants

Twenty-eight orthopedic residents in four consecutive residency years (year 3 to 6) and ten experienced orthopedic surgeons were recruited at two institutions (the Stanford University School of Medicine and the Academic Medical Centre in Amsterdam) (Table 1).

#### Study design

After participants signed informed consent, all outpatient

---

<table>
<thead>
<tr>
<th>Number</th>
<th>Description</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Dissection</td>
<td>1: Appeared excessively hesitant, caused trauma to tissues, did not dissect into correct anatomical plane. 2: Controlled and safe dissection into correct anatomical plane, caused minimal trauma to tissues. 3: Superior and atraumatic dissection into the correct anatomical plane.</td>
</tr>
<tr>
<td>2</td>
<td>Instrument handling</td>
<td>1: Repeatedly makes tentative or awkward movements with instruments. 2: Competent use of instruments, although occasionally appeared stuff or awkward. 3: Fluid moves with instruments and no awkwardness.</td>
</tr>
<tr>
<td>3</td>
<td>Depth perception</td>
<td>1: Constantly overshoots target, slow to correct. 2: Some overshooting or missing of target. 3: Accurately directs instruments in the correct plane to target.</td>
</tr>
<tr>
<td>4</td>
<td>Bimanual dexterity</td>
<td>1: Noticeably awkward with non-dominant hand, poor coordination between hands. 2: Uses both hands but does not maximize interaction between hands. 3: Expertly uses both hands in complementary manner to provide optimum performance.</td>
</tr>
<tr>
<td>5</td>
<td>Flow of operation and forward planning</td>
<td>1: Frequently stopped operating or needed to discuss next move. 2: Demonstrated ability for forward planning with steady progression of operative procedure. 3: Obviously planned course of operation with effortless flow from one move to the next.</td>
</tr>
<tr>
<td>6</td>
<td>Knowledge of instruments</td>
<td>1: Frequently asked for the wrong instrument or used inappropriate instrument. 2: Knew the names of most instruments and used appropriate instrument for the task. 3: Obviously familiar with the instruments required and their names.</td>
</tr>
<tr>
<td>7</td>
<td>Efficiency</td>
<td>1: Many unnecessary, inefficient movements. Constantly changing focus or persisting without progress. 2: Slow, but planned movements are reasonably organized with few unnecessary or repetitive movements. 3: Confident, clear economy of movement and maximum efficiency.</td>
</tr>
<tr>
<td>8</td>
<td>Knowledge of specific procedure</td>
<td>1: Deficient knowledge, needed specific instruction at most operative steps. 2: Knew all important aspects of the operation. 3: Demonstrated familiarity with all aspects of the operation.</td>
</tr>
<tr>
<td>9</td>
<td>Autonomy</td>
<td>1: Unable to complete entire task, even with verbal guidance. 2: 3: Able to complete task safely with moderate guidance. 4: 5: Able to complete task independently</td>
</tr>
<tr>
<td>10</td>
<td>Quality of final product</td>
<td>1: Very poor. 2: Competent. 3: Clearly superior.</td>
</tr>
</tbody>
</table>

---

*Figure 1: GRS I Basic Arthroscopic Knee Skill Scoring System (BAKSSS) [16]*
KA's and ACL reconstructions performed by each resident were supervised and assessed by one of the experienced surgeons, using both GRS questionnaires. Before each procedure, the resident’s and supervisor’s unique identifier code, type of operation, year of residency and number of previously performed arthroscopies were documented. Hundred-and-thirty-five KA's and 30 ACL reconstructions were included (Table 1). Residents performed on average six procedures, within a time frame of maximum a month.

### Outcome measures

GRS I is a ten-item Global Rating Scale, derived from previously published and validated evaluation models to assess arthroscopic skills (Figure 1). Items from these models were used to create a task-specific checklist and a global rating scale that together form the Basic Arthroscopic Knee Skill Scoring System (BAKSSS), a model specific to diagnostic knee arthroscopy and partial meniscectomy [7]. The checklist was not included in the current study. Items of the global rating scales include dissection, instrument handling, depth perception, bimanual dexterity, flow of operation, knowledge of instruments, to the knowledge of the specific procedure, autonomy, efficiency, and quality of the operative result [7].

For GRS II, nine of the fourteen Orthopaedic Competence Assessment Project (OCAP) criteria for diagnostic arthroscopy were selected (Figure 2). Competences include following procedure protocol, handling of tissue, appropriate and safe use of instruments, appropriate pace with economy movement, calmness and effectiveness in dealing with untoward events, appropriate use of assistants, communication with scrub nurse, and identification of common abnormalities and protection of articular surface [24].

GRS I and II have similar domains, such as instrument handling, flow of operation, efficiency and autonomy. Both Global Rating Scales allow assessors to rate arthroscopic skill performance on each domain, using 5-point Likert scales with anchors at 1, 3, and 5 points. The anchor points have specific descriptions of the necessary requirements to receive the respective point values, which should help uniform assessment. Higher scores indicate better arthroscopic proficiency. Minimum and maximum GRS score for the GRS I are 10 and 50 points, respectively [7]. Minimum and maximum GRS score for GRS II are 9 and 45 points, respectively [24].

Firstly, known-groups validity will be investigated by determining the extent to which the GRSs can discriminate between levels of experience [25]. Secondly, convergent validity is investigated by determining whether the two GRSs correspond with one another, as they cover similar domains of arthroscopic skills [26]. To this end, knee arthroscopies (KA) and anterior cruciate ligament (ACL) reconstructions performed on real life patients will be used. If validity is shown, these GRS could be further developed into objective assessment tools to show individual training progress of residents.

### Statistical analysis

Analysis was performed with SPSS 22 © (SPSS Inc., Chicago, IL, USA). All 165 procedures were included in the analysis (Table 1). In order to compare GRS I and II, scores were normalized to a range from 0 to 100 points. Normality of the parameters was assessed using the Kolmogorov-Smirnov test, and the skewness and kurtosis of the sample (-2 < z-value < 2).

To account for correlated assessments within the residents (that is multiple assessments were performed per resident), a multilevel analysis was performed by use of mixed model analysis using a residual maximum likelihood (REML) approach (Table 2). The estimate of the parameters, as well as the standard error and confidence intervals were used to study the relationship between year of residency and each GRS score, and between the number of previously performed arthroscopies and each GRS score. Type of operation (KA of ACL reconstruction), year of residency and number of previously performed arthroscopies were entered as model factors, with GRS I and GRS II scores as dependent variables. Pearson correlation coefficient between the scores of GRS I and GRS II were calculated to measure convergent validity. Correlation coefficients ≤ 0.35 were considered to represent weak correlations, 0.36 to 0.67 moderate correlations, and 0.68 to 1.0 high correlations, with coefficients ≥ 0.91 very high correlations [27,28].

A Bland-Altman plot was constructed to evaluate the agreement between GRS I and II. The mean differences between GRS I and II scores against the absolute differences and limits of agreement (LoA) were calculated (1.96*SDpu) [29]. A paired t-test was performed to assess a systematic difference between the two scales. P-values ≤ 0.05 were considered statistically significant.

### Results

Non-normalized GRS sum scores varied between 19 and 50 points for GRS I and between 18 and 45 points for GRS II. In Table 2, results of the mixed model analysis are described. The parameters can be interpreted as the constant (intercept) and the coefficients or slopes (estimates) of the independent variables. Mixed model analysis revealed a statistically significant increase (p < 0.001) per year of residency on both GRS I and II, with values of 8.1 points (95% CI: 6.3-9.9) and 9.2 points (95% CI: 7.4-11.2) respectively. Significant increases per performed arthroscopy were also observed for both GRSs (p < 0.001), with values of 0.14 (95% CI: 0.09-0.18) and 0.13 points (95% CI: 0.10-0.1) for GRS I and II respectively. Furthermore, scores for ACL reconstructions were significantly lower (p < 0.001) than for standard knee arthroscopies (12.5 and 13.0 points for GRS I and II respectively, p < 0.001) (Table 2).

Normalized GRS sum scores varied between 40 and 100 points for GRS I and between 38 and 100 points for GRS II. The scores did not differ significantly (p = 0.19), with mean normalized scores of 70.8 (SD is 14.9) for GRS I and 71.3 (SD is 16.2). The Pearson correlation coefficient between the normalized GRS I and GRS II scores was high (0.94). The calculated LoA was 11 points, resulting in a lower limit of -11.6 and an upper limit of 10.4 (Figure 3).

### Discussion

This study investigated if the proposed GRSs show construct validity, more specifically, known-groups validity and convergent validity. With the available sample size, the study demonstrated that both GRS I and II were able to discriminate based on year of residency or number of arthroscopies, supporting known-groups validity.

<table>
<thead>
<tr>
<th>Year of Residency</th>
<th>No. of residents</th>
<th>Total no. of procedures</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>8</td>
<td>32</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>36</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>45</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>52</td>
</tr>
</tbody>
</table>

| Table 2: Mixed model analysis showing the effects of year of residency, number of previously performed procedures and type of operation on scores for GRS I and II. All estimates are significant (p < 0.001). The intercept can be interpreted as the mean of the outcome when all independent variables are zero. The estimates can be interpreted the same way as the estimates (coefficients) of predictors in a linear regression. GRS scores increase with 8.1 and 9.3 respectively per year of residency, and with 0.14 and 0.13 points respectively per number of previously performed arthroscopies. Scores for ACL-procedures are on average 12.5 and 13.0 points lower that KA scores. |
|--------------------------|--------------------------|--------------------------|
| GRS I | GRS II | Estimate (95% CI) | Estimate (95% CI) |
| Intercept | 34.4 (25.7-43.1) | 29.4 (20.1-38.7) |
| Year of residency | 8.1 (6.3-9.4) | 9.3 (7.4-11.2) |
| No of arthroscopies | 0.14 (0.1-0.2) | 0.13 (0.1-0.2) |
| ACL-procedure | -12.5 (-17.2--7.8) | -13.0 (-17.8--7.8) |

s.e.: standard error; CI: confidence interval.
Convergent validity of the studied Global Rating Scales was supported by a high Pearson correlation coefficient. The Bland and Altman plot demonstrated that the average discrepancy between GRS I and II was small (close to zero), indicating that there was no systematic difference between the two scales. However, the limits of agreement had higher values than the estimated differences per residency year or arthroscopic intervention: the standard error was 11, meaning that only differences larger than 11 points between the GRSs can be interpreted as actual difference when used on individual level. This emphasizes the question whether the GRSs are reliable outcome measures and if they are suitable for performance monitoring on individual level.

Besides year of residency, GRS scores are influenced by other factors, which can account for variability. One of these factors is the complexity of the type of procedure: ACL reconstructions are more complex than knee arthroscopies, which was reflected in the significantly lower GRS scores for ACL reconstructions. Other factors influencing GRS scores are the complexity of the joint (depending on the anatomy or the severity of the condition of the patient), inter-observer differences between supervisors.

### Figure 2: GRS II Orthopaedic Competence Assessment Project (OCAP) [14]
and the moment of the day or of the week at which the procedure was performed. Thus, as GRS score is determined by other factors additional to level of experience, in particular type and complexity of procedure, a standardized setting is required when using the GRS to measure competence.

The GRSs did not show floor or ceiling effects. None of the participants scored lower than 19 (GRS I) and 18 (GRS II) points, whereas the minimum values of the scales are 10 and 9, respectively. Moreover, no item was scored below two points. This can be attributed to the range of residency years that was included: residents were selected from their third year of residency, as they start than with their specialisation in orthopaedic surgery in the Netherlands. Hence, none of the residents participating in the current study was a completely untrained novice, as opposed to the original studies that also included participants with none or very little training [7,24]. Our results indicate that the GRSs can be used for the entire duration of the residency curriculum.

The current study has limitations. Firstly, residents were assessed by one supervisor, implicating that inter-observer reliability could not be assessed. Secondly, supervisors were not blinded and thus were aware of the level of training of the residents. Thirdly, supervisors were not specifically trained to use the two Global Rating Scales of procedure. Ideally, all supervisors should have been trained with distinct examples, and multiple observers, who were blinded from the identity of the residents, should have performed the scoring. This would have increased consensus and objectivity. Unfortunately, this was logistically difficult to arrange. Vogt showed that knowing the identity of the resident does not significantly affect scoring [30]. Moreover, the study design was similar to other studies showing the potential of Global Rating Scales to objectively evaluate arthroscopic skills [7,24,31]. Therefore, we expect that these limitations will have marginally influenced the results.

As competency-based education is becoming more important in arthroscopic training [32], objective tools for assessment and performance monitoring of orthopedic residents need to be validated. The current study showed known-groups validity and convergent validity for Global Rating Scales. However, the results also suggest the scales do not seem to be sufficiently sensitive and consistent to monitor individual learning curves and progress of a trainee over a short period of time. Rather, they are suitable to objectify and assess general arthroscopic performance on group level in a structured way. Moreover, they can be applied in a research setting; to study differences on group level and to perform sample size calculation required to detect significant differences between different levels of experience. Lastly, as feedback on performance is known to improve the learning process [33-36], and the structure of the GRSs allows feedback per skill domain, GRSs can also be valuable as educational tools.

Conclusion

The Basic Arthroscopic Knee Skill Scoring System (GRS I) and the Orthopaedic Competence Assessment Project (GRS II) demonstrate sufficient construct validity when performing knee arthroscopy or ACL reconstruction. However, they seem not sufficiently sensitive and consistent to establish individual learning curves. Both scales are suitable to objectively evaluate global progress of residents in the operating room when acquiring arthroscopic skills, in particular on group level.

Acknowledgement

The authors are grateful to Joris Landaal and Mikel Reilingh for their contribution to enable the GRS assessments, and Robert van de Breek for his contribution to the set up and the execution of this study. We also want to thank the Slotervaart Hospital in Amsterdam, the Tergooi Hospital in Hilversum for their cooperation.

References