



## City Research Online

### City, University of London Institutional Repository

---

**Citation:** Vianya-Estopa, M., Nagra, M., Cochrane, A., Retallic, N., Dunning, D., Terry, L., Lloyd, A., Wolffsohn, J. S., Alderson, A., Mouat, G., et al (2020). Optimising subjective anterior eye grading precision. *Contact Lens and Anterior Eye*, 43(5), pp. 489-492. doi: 10.1016/j.clae.2020.03.006

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

---

**Permanent repository link:** <https://openaccess.city.ac.uk/id/eprint/24191/>

**Link to published version:** <https://doi.org/10.1016/j.clae.2020.03.006>

**Copyright:** City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

**Reuse:** Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

---

City Research Online:

<http://openaccess.city.ac.uk/>

[publications@city.ac.uk](mailto:publications@city.ac.uk)

---

## **Title: Optimising Subjective Anterior Eye Grading Precision**

<sup>1</sup>Marta Vianya-Estopa PhD [marta.vianya@anglia.ac.uk](mailto:marta.vianya@anglia.ac.uk)

<sup>2</sup>Manbir Nagra PhD [manbir.nagra@port.ac.uk](mailto:manbir.nagra@port.ac.uk)

<sup>3</sup>Arnold Cochrane BSc [ajs.cochrane@ulster.ac.uk](mailto:ajs.cochrane@ulster.ac.uk)

<sup>4</sup>Neil Retallic BSc [optomneil@outlook.com](mailto:optomneil@outlook.com)

<sup>5</sup>Dean Dunning MEd [D.Dunning@bradfordcollege.ac.uk](mailto:D.Dunning@bradfordcollege.ac.uk)

<sup>6</sup>Louise Terry PhD [TerryL1@cardiff.ac.uk](mailto:TerryL1@cardiff.ac.uk)

<sup>7</sup>Aoife Lloyd PhD [aoife.lloyd@mckernan@dit.ie](mailto:aoife.lloyd@mckernan@dit.ie)

<sup>8</sup>James S Wolffsohn PhD [j.s.w.wolffsohn@aston.ac.uk](mailto:j.s.w.wolffsohn@aston.ac.uk)

and members of the British & Irish University & College Contact Lens Educators (BUCCLE)

### Affiliations

<sup>1</sup>Vision and Hearing Sciences, Anglia Ruskin University, Cambridge, UK

<sup>2</sup>Optometry, University of Portsmouth, Portsmouth, UK

<sup>3</sup>School of Biomedical Sciences, Ulster University, Coleraine, UK

<sup>4</sup>Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, UK

<sup>5</sup>Advanced Technology Centre, Bradford College, Bradford, UK

<sup>6</sup>School of Optometry and Vision Sciences, Cardiff University, Cardiff, UK

<sup>7</sup>School of Physics & Clinical & Optometric Sciences, Technological University Dublin, Ireland.

<sup>8</sup>Ophthalmic Research Group, Aston University, Birmingham, UK

**Corresponding author: Prof James S Wolffsohn, Aston University, Aston Triangle, Birmingham, B4 7ET, UK [j.s.w.wolffsohn@aston.ac.uk](mailto:j.s.w.wolffsohn@aston.ac.uk)**

Members of the British & Irish University & College Contact Lens Educators (BUCCLE) are: Alison Alderson and Graham Mouat (University of Bradford), Claire McDonnell and Orla Murphy (TU Dublin), Byki Huntjens and Mark Mayhew (City, University of London), Eilidh Martin (Glasgow Caledonian University), Laura Sweeney (Glasgow Caledonian University), Katherine Evans (Cardiff University), Shehzad Naroo (Aston University), Robert Conway (Anglia Ruskin University), Luisa Simo (Plymouth University), Carole Maldonado-Codina and Claire Mallon (The University of Manchester), Jo Underwood (Association of British Dispensing Opticians), Kishan Trivedy (University of Portsmouth) and Mahesh Joshi (University of Plymouth).

## **Abstract**

**Purpose:** To establish the optimum grading increment which ensured parity between practitioners while maximising clinical precision.

**Methods:** Second year optometry students (n=127, 19.5 ± 1.4 years, 55% female) and qualified eye care practitioners (n=61, 40.2 ±14.8 years, 52% female) had 30 seconds to grade each of bulbar, limbal and palpebral hyperaemia of the upper lid of 4 patients imaged live with a digital slit lamp under 16x magnification, diffuse illumination, with the image projected on a screen. The patients were presented in a randomised sequence 3 times in succession, during which the graders used the Efron printed grading scale once to 0.1 precision, once to 0.5 precision and once to the nearest integer grade in a randomised order. Graders were masked to their previous responses.

**Results:** For most grading conditions less than 20% of clinicians showed a  $\leq 0.1$  difference in grade from the mean. In contrast, more than 50% of the student graders and 40% of experienced graders showed a difference in grade from the mean within 0.5 for all conditions under measurement. Student precision in grading was better with both 0.1 and 0.5 grading precision than grading to the nearest unit, except for limbal hyperaemia where they performed more accurately with 0.5 unit precision grading. Limbal grading precision was not affected by grading step precision for experienced practitioners, but 0.1 and 0.5 grading precision were both better than 1.0 grading precision for bulbar hyperaemia and 0.1 grading precision was better than 0.5 grading precision and both were better than 1.0 grading precision for palpebral hyperaemia.

**Conclusion:** Although narrower intervals scales maximise the ability to detect smaller clinical changes, the grading increment should not exceed one standard deviation of the discrepancy between measurements. Therefore, 0.5 grading increments are recommended for subjective anterior eye physiology grading (limbal, bulbar and palpebral redness).

**Keywords:** grading; hyperaemia; student; eye-care practitioner; scale increments

## 1 Introduction

2 Since their initial introduction approximately thirty years ago, anterior eye grading scales have firmly  
3 established themselves as an essential part of the eye care practitioner's (ECPs) armamentarium.  
4 With usage reported at approximately 60-85% amongst ECPs [1,2] this seemingly low-tech approach  
5 has had a significant impact on clinical practice. Grading scales hold several advantages over the  
6 sole use of written descriptions and sketches that practitioners had previously relied upon. Grading  
7 scales are quantitative, simplify the monitoring and progression of pathological and physiological  
8 changes, are a universal familiar language so can be interpreted by different nationalities and across  
9 health care professionals, aid in medical legal cases, and ultimately facilitate patient management.

10 While grading scales are easy to use, widely available, and considered best practice [2], they are not  
11 without their limitations. Grading is subjective, associated with poor repeatability [3] and high  
12 variability amongst practitioners. Grading scales are not interchangeable and the scale range varies,  
13 thus grading scores will differ depending on scale used [4] with estimates reported to be higher for  
14 scales which have a shorter dynamic range. [5] Further, there are concerns about the grading  
15 reference images themselves. Wolffsohn [6] found grading scale images did not follow a linear  
16 increase in severity, but instead followed a quadratic pattern, such that precision is greater for lower  
17 severity reference images i.e. the increments between gradings are unequal. Digital versions of  
18 grading scales have been produced with morphing technology [7] used to generate reference images  
19 down to 0.1 scale grade increments, but any improvement in grading variability has not been  
20 published.

21 Some of the shortcomings may be attributable to the process of grading itself; typically, anterior eye  
22 grading involves the application of a discrete scale (a limited fixed number of grades) to a continuous  
23 variable (the severity of a particular ocular condition). [8] Several sources [2,8] have advocated the  
24 reduction of grading scale increment size to increase clinical precision i.e. grading to the nearest  
25 integer should produce poorer clinical precision than grading to the nearest 0.5 or 0.1.

26 Nevertheless, achieving adequate clinical precision may not necessitate use of the smallest grading  
27 increment possible. Peterson and Wolffsohn [3] showed a mean difference of approximately 0.70-  
28 1.03 bulbar redness (Efron) image grades was needed for it to be discernible by subjective grading.  
29 Given the widespread use of grading scales, and their vulnerability to subjective bias, it is of clinical  
30 interest to establish an evidence base for a best practice approach to grading. The aim of this study  
31 was to establish the optimum grading increment which ensured parity between practitioners while  
32 maximising clinical precision. Based on previously published data, it is hypothesised that whole  
33 integer grading will be less accurate (a larger absolute deviation from the mean practitioner grade)  
34 than grading to the nearest 0.5 or 0.1 unit.

## 35 Method

36 The study was granted a favourable ethical opinion by Ulster University (practitioner study) and  
37 Aston University (student study) ethics committees and followed the tenets of the Declaration of  
38 Helsinki. Participants gave written informed consent to take part after an explanation of the study.

39 The graders were 2<sup>nd</sup> year undergraduate optometry students enrolled at Aston University (n=127,  
40 19.5 ± 1.4 years, 55% female) and qualified eye care practitioners (at least 2 years) attending the  
41 BCLA UK conference in June 2018 (n=61, 40.2 ±14.8 years, 52% female) all familiar with using grading  
42 scales with the Efron grading scale. Data collection for the two cohorts occurred on separate  
43 occasions.

44 The ocular surface of 4 patients with no ocular pathology were observed live under 16x  
45 magnification, diffuse illumination, with a digital slit-lamp (Keeler, Windsor, UK) and the image  
46 projected on a screen. The patients were presented in a randomised sequence 3 times in succession  
47 during which the graders used the Efron printed grading scale once to 0.1 increments, once to 0.5  
48 increments and once to the nearest integer grade in randomised order. They had 30 seconds to  
49 grade each of bulbar, limbal and upper lid palpebral hyperaemia each time, and were masked to  
50 their previous grades.

#### 51 Statistical Analysis

52 The absolute average difference from the mean of all graders, for each grader with each increment  
53 level was calculated for each of the 4 patients. As the data was not normally distributed, non-  
54 parametric statistics were applied (Friedman test repeated measure analysis of variance with  
55 Wilcoxon signed-rank test post-hoc pairwise comparison where significance was identified). In  
56 addition the discrepancies between pairs of observers were assessed for each of the 4 patients and  
57 the standard deviation calculated.

58 Based on a standard deviation of 0.4 [9] for subjective grading, a clinically significant difference  
59 ( $p < 0.05$ ) of 0.2 units between groups could be detected with 80% power with a sample size of 61  
60 participants in each group and 0.15 units with 127 participants in each group.

61 <https://www.stat.ubc.ca/~rollin/stats/ssize/n2.html>

62

63 **Results**

64 Across the 4 patients examined, the average bulbar grade ranged from 0.8-1.5, average limbal grade  
 65 ranged from 0.4 to 1.2 and the average palpebral grade ranged from 0.4 to 1.6 and was similar  
 66 between patients used for the student grading and practitioner grading sessions. The distribution of  
 67 the difference from the mean is shown in Figure 1 for student graders and Figure 2 for qualified eye  
 68 care practitioners. The mean of these differences for each feature is shown in Table 1, along with  
 69 statistical significance. There was a significant difference ( $p < 0.001$ ) across all grade increment  
 70 comparisons except practitioner graded limbal hyperaemia ( $p = 0.478$ ). The percentage of clinicians  
 71 increased for all conditions within a greater difference in grade from the mean. For most conditions  
 72 less than 20% of clinicians showed a  $\leq 0.1$  difference in grade from the mean. In contrast, more than  
 73 50% of the student graders and 40% of experienced graders showed a difference in grade from the  
 74 mean within 0.5 for all conditions under measurement.

Grading		0.1		↔	0.5		↔	1.0		↔ 0.1
Increment		mean	p		mean	p		Mean	P	
Student n=127	Bulbar	0.40±0.30	0.342		0.42±0.31	<b>&lt;0.001</b>		0.51±0.29	<b>&lt;0.001</b>	
	Limbal	0.44±0.31	0.156		0.42±0.31	<b>0.001</b>		0.47±0.36	0.259	
	Palpebral	0.35±0.26	0.645		0.34±0.32	<b>&lt;0.001</b>		0.49±0.27	<b>&lt;0.001</b>	
Practitioner n=61	Bulbar	0.58±0.50	0.633		0.58±0.53	<b>0.004</b>		0.64±0.45	<b>0.001</b>	
	Limbal	0.54±0.46	0.790		0.54±0.49	0.940		0.53±0.52	0.874	
	Palpebral	0.71±0.64	<b>0.026</b>		0.75±0.67	<b>&lt;0.001</b>		0.82±0.64	<b>&lt;0.001</b>	

75 **Table 1:** Mean grade difference ( $\pm$  S.D.) from mean and significance between grading  
 76 increments. The arrows above the significance (p) values point to the two  
 77 increments being compared.

78

79 Student precision in grading was better with both 0.1 and 0.5 grading increments than grading to the  
 80 nearest unit, except for limbal hyperaemia where it was only better with 0.5 unit increment grading  
 81 (there was no significant difference between the 0.1 and 1.0 increments for this feature). Limbal  
 82 grading precision was not affected by grading step increment for experienced practitioners, but 0.1  
 83 and 0.5 grading increments were both better than the 1.0 grading increment for bulbar hyperaemia.  
 84 For palpebral hyperaemia, the 0.1 grading increment was more accurate than the 0.5 grading  
 85 increment and both were better than 1.0 grading increment (Table 1). The standard deviation of  
 86 discrepancies between observers was 0.65-0.87 across the students and was 0.72 to 0.84 across  
 87 experienced practitioners.

88

89 **Figure 1:** What proportion of student clinicians were within 0.1 to 1.0 grades different from the  
90 mean of all clinicians for bulbar, limbal, and palpebral hyperaemia with each of the  
91 grading increments. N=127.

92 **Figure 2:** What proportion of experienced practitioners within 0.1 to 1.0 grades different from  
93 the mean of all clinicians for bulbar, limbal and palpebral hyperaemia with each of the  
94 grading increments. N=61.

95

## 96 **Discussion**

97 This study set out to show that smaller grading increment steps would lead to more accurate grading  
98 compared to the mean. In practical terms, the grades recorded by a practitioner should be as close  
99 as possible to the mean of other practitioners (average difference) rather than the discrepancy  
100 analysis (the difference between 2 practitioners) as modelled by Bailey et al. [8]. However, while this  
101 was the case for 0.5 grading units compared to whole integer grading, this was generally not the  
102 case for 0.1 grading units compared to 0.5. As shown in Table 1, the average difference from the  
103 mean was around 0.30 for student graders and 0.55 for experienced graders. The standard deviation  
104 between random pairs of observers was higher, as expected, being 0.72 for student graders and 0.78  
105 for experienced graders. Bailey et al. [8] suggested that if the scale increment exceeds the standard  
106 deviation of the discrepancy this will result in a sharp broadening of the confidence limits. Thus,  
107 these findings suggest that a 0.5 grading step might be as precise as is possible to get when  
108 evaluating hyperaemia in the anterior eye using the Efron printed grading scale.

109 It is worth noting that limbal hyperaemia grading was more variable in grade than bulbar and  
110 palpebral redness. This finding is not surprising as the exact extent of the limbal region is not clearly  
111 defined clinically and graders might have been influenced by nearby conjunctival redness. Yet,  
112 observers need to ensure enough attention is given to this structure given the response between  
113 limbal hyperaemia and contact lens wear. For instance, several studies have shown that hydrogel  
114 lens wear results in significantly greater levels of limbal hyperaemia compared to silicone hydrogel  
115 lens wear for both daily and extended wear modalities, whereas bulbar redness is not significantly  
116 affected. [10-13]

117 Efron et al [4] suggested that grading of contact lens complications would be expected to improve  
118 with experience. His group also found grading variability improves statistically (but not clinically  
119 significant) with some experience, but no added benefit could be derived from supplemental  
120 training [14]. However, this study found experienced practitioners were less accurate than second  
121 year undergraduate optometry students. Similar findings between students and experienced  
122 practitioners were also noted by Wolffsohn et al [4]. Although a priori one might expect experienced



123 practitioners to show greater precision than students, this might no longer be the case as the  
124 importance of grading in the assessment of anterior eye is currently emphasised to undergraduate  
125 optometry students. Similarly, Cardona and Serés [15] noted that contact lens knowledge improved  
126 grading precision in optometry students. The students taking part in this study had received a 1 hour  
127 seminar on the principals behind grading and had used the Efron grading scales in 5 weekly 2 hour  
128 clinics. Differences in the data projection of the images, such as screen resolution and ambient  
129 brightness could have made a difference between cohorts, but the student graders used the same  
130 conditions as half of the experienced graders and the difference between them was still evident.  
131 Future work should further explore the relationship that knowledge, training and experience might  
132 have on the uses of grading scales in anterior eye and contact lens assessment. A survey of UK  
133 practitioners in 2015 [2] indicated that 91.6% of respondents used grading scales for bulbar  
134 conjunctival hyperaemia and 77.8% and 63.4% for limbal and palpebral hyperaemia respectively. It  
135 could be hypothesised that less familiarity of usage might lead to more variability with grading and  
136 this seems to be the case with practitioners.

137 Recently, alternative methods to subjective assessment of bulbar and limbal hyperaemia have been  
138 proposed using software such as Keratograph 5M (Oculus) that objectively detects hyperaemia. [16]  
139 Artificial intelligence learning algorithms have been applied to retinal images, demonstrating their  
140 ability not just to quantify disease changes, but also to identify other features that might  
141 differentiate disease and its progression such as tortuosity, pallor and blood flow, not traditionally  
142 utilised by clinicians. [17] However, technological advances are not yet readily available by most  
143 clinicians. In addition, the results of this new technology might not be interchangeable with results  
144 obtained using subjective grading scales. [18-19] Thus, it is important to continue to support  
145 clinicians using grading scales optimally, although, digital photography can allow direct comparison  
146 at subsequent visits and is preferable to grading.

147 It is important to note that this study was conducted using projected slit lamp videos of eyes without  
148 pathology. The patients examined were different between the students and experienced  
149 practitioners, but the average grades were similar for each of the ocular anterior eye features  
150 examined and the comparison was the individual's difference from the mean, so the actual mean  
151 should not have a significant effect on the results. The mean grade of each feature was  $\leq 2$  for each  
152 participant; the entire range of the grading scale used was not included in the study. Therefore the  
153 conclusions cannot be extended to grading precision for more severe hyperaemic cases.

154 In conclusion, this study showed that 0.5 grading increments should be recommended when  
155 assessing anterior eye grading (limbal, bulbar and palpebral hyperaemia). This contradicts previous  
156 recommendation by Efron et al [4] and Wolffsohn et al. [2] of recording clinical signs using 0.1  
157 increments between grades. Although narrower intervals scales maximise the ability to detect  
158 smaller clinical changes, Bailey et al [8] also indicated that for moderate precision the grading

159 increment should not exceed one standard deviation of the discrepancy between measurements.  
160 Although narrower increments have been recommended in clinical practice, Efron et al [4] and  
161 Wolffsohn et al [2] found graders tended to grade using whole and half-digits indicating a reluctance  
162 to use finer increments. Thus, this research provides the evidence for clinicians to adopt 0.5  
163 increments in their clinical grading alongside previous research highlighting the importance of  
164 recording the scale used and having the scale present when grading. [2,6]

165

166

167 **References**

- 168 [1] Efron N, Pritchard N, Brandon K, Copeland J, Godfrey R, Hamlyn B, Vrbancic V. A survey of the use  
169 of grading scales for contact lens complications in optometric practice. *Clin Exp Optom* 2011;94:193-  
170 9.
- 171 [2] Wolffsohn JS, Naroo SA, Christie C, Morris J, Conway R, Maldonado-Codina C. Anterior eye health  
172 recording. *Contact Lens Ant Eye* 2015;38:266-71.
- 173 [3] Peterson RC, Wolffsohn JS. Sensitivity and reliability of objective image analysis compared to  
174 subjective grading of bulbar hyperaemia. *Br J Ophthalmol* 2007;91:1464-6.
- 175 [4] Efron N, Morgan PB, Katsara SS. Validation of grading scales for contact lens complications.  
176 *Ophthalmic Physiol Opt.* 2001;21:17-29.
- 177 [5] Schulze MM, Hutchings N, Simpson TL. Grading bulbar redness using cross-calibrated clinical  
178 grading scales. *Invest Ophthalmol Vis Sci* 2011;52:5812-7.
- 179 [6] Wolffsohn JS. Incremental nature of anterior eye grading scales determined by objective image  
180 analysis. *Br J Ophthalmol* 2004; 88:1434-8.
- 181 [7] Efron N, Morgan PB, Jagpal R. Validation of computer morphs for grading contact lens  
182 complications. *Ophthalmic Physiol Opt* 2002;22:341-49.
- 183 [8] Bailey IL, Bullimore MA, Raasch TW and Taylor HR. Clinical Grading and the Effects of Scaling.  
184 *Invest Ophthalmol Vis Sci* 1991;32:422–32.
- 185 [9] Efron N. Grading scales for contact lens complications. *Ophthalmic Physiol Opt* 1998;18:182-6.
- 186 [10] Brennan NA, Coles ML, Connor HR, McIlroy RG. A 12-month prospective clinical trial of  
187 comfilcon A silicone-hydrogel contact lenses worn on a 30-day continuous wear basis. *Cont Lens*  
188 *Anterior Eye* 2007;30:108-18.
- 189 [11] Dumbleton K, Keir N, Moezzi A, Feng Y, Jones L, Fonn D. Objective and subjective responses in  
190 patients refitted to daily-wear silicone hydrogel contact lenses. *Optom Vis Sci* 2006; 83:758-68.
- 191 [12] Maldonado-Codina C, Morgan PB, Schnider CM, Efron N. Short-term physiologic response in  
192 neophyte subjects fitted with hydrogel and silicone hydrogel contact lenses. *Optom Vis Sci* 2004;  
193 81:911-21.
- 194 [13] Dillehay SM, Miller MB. Performance of Lotrafilcon B silicone hydrogel contact lenses in  
195 experienced low-Dk/t daily lens wearers. *Eye Contact Lens* 2007; 33:272-77.

- 196 [14] Efron N, Morgan PB, Farmer C, Furuborg J, Struk R, Carney LG. Experience and training as  
197 determinants of grading reliability when assessing the severity of contact lens complications.  
198 *Ophthalmic Physiol Opt* 2003;23:119-24.
- 199 [15] Cardona G & Serés C. Grading Contact Lens Complications: The Effect of Knowledge on Grading  
200 Accuracy. *Curr Eye Res* 2009; 34: 1074–81.
- 201 [16] Wu S, Hong J, Tian L, Cui X, Sun X and Xu J. Assessment of Bulbar Redness with a Newly  
202 Developed Keratograph. *Optom Vis Sci* 2015; 92: 892-899.
- 203 [17] Varadarajan AV, Poplin R, Blumer K, Angermueller CA, Ledsam J, Chopra R, Keane PA, Corrado  
204 GS, Peng L and Webster DR. Deep learning for predicting refractive error from retinal fundus images.  
205 *Invest Ophthalmol Vis Sci* 2018;59: 2861-8.
- 206 [18] Perez-Bartolome F and Garcia-Feijoo J. Assessment of ocular redness measurements obtained  
207 with keratography 5M and correlation with subjective grading scales. *Journal Français*  
208 *d’Ophthalmologie* 2018; 41 (9): 836-46.
- 209 [19] Huntjens B, Basi M, Nagra M. Evaluating a new objective grading software for conjunctival  
210 hyperaemia. 2019. *Contact Lens Ant Eye* (in press)

211

## 212 **Acknowledgements**

213 BUCCLE’s mission is to enhance optometry education in the UK and in this pursuit is funded by  
214 industry including Alcon, Bausch and Lomb, Coopervision, No7 Contact lenses and David  
215 Thomas/Menicon. This work was also supported by Cheryl Donnelly and the BCLA. Other BUCCLE  
216 members include Alison Alderson and Graham Mouat (University of Bradford), Claire McDonnell and  
217 Orla Murphy (TU Dublin), Byki Huntjens (City, University of London), Mark Mayhem (City, University  
218 of London), Eilidh Martin (Glasgow Caledonian University), Laura Sweeney (Glasgow Caledonian  
219 University), Katherine Evans (Cardiff University), Shehzad Naroo (Aston University), Robert Conway  
220 (Anglia Ruskin University), Luisa Simo (Plymouth University), Carole Maldonado-Codina and Claire  
221 Mallon (The University of Manchester), Jo Underwood (Association of British Dispensing Opticians),  
222 Kishan Trivedy (University of Portsmouth) and Mahesh Joshi (University of Plymouth).