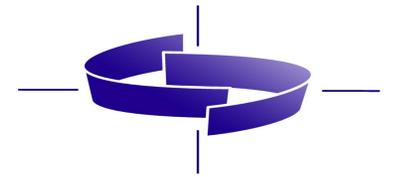


Comparative study of two grading scales for the evaluation of IR-meibography images of the upper and lower eyelid in terms of intra- and interrater reliability

B.Sc. Anke von Ahrentschildt¹; M.Sc. Julia Wittekind²; Dipl.-Ing. S. Marx²; Prof., M.Sc., Dipl.-Ing. W. Sickenberger^{1,2}
¹Ernst Abbe University of Applied Sciences Jena, Germany, ²JENVIS Research Jena



Introduction:

- In **11.1 %** dry eye symptoms are caused by a hyopsecretion of the lacrimal gland. More often hyperevaporation is the initiating factor with **76.7 %** [1].
- A dysfunction of the **meibomian glands** induces an insufficient lipidlayer which leads to a hyperevaporation of the tear film [2]



Therefore the examination of the meibomian glands with infrared-meibography is supposed to be a standard component of the dry eye diagnostic

- IR-meibography uses IR-light and an IR-sensitive camera to visualize the meibomian glands → signs of a dysfunction can be detected at an early stage
- Grading scales enable a meaningful documentation
- For the classification of meibomian glands the MEIBOSCORE by ARITA et al. [2] seems to be the most common grading scale → lacking information regarding its quality

Purpose:

Primary objective:

- Intra- and interrater reliability of the new illustrative grading scale by JENVIS RESEARCH which was developed in dependence on the MEIBOSCORE of ARITA et al. (2008)

Secondary objective:

- Intra- and interrater reliability for the MEIBOSCORE
- Compare both scales with each other

Methods:

- 48 trained practitioners independently graded a set of 80 images, 40 of the upper as well as the lower eyelid in a randomized order, with one of two grading scales in an online survey that consisted of four independent cycles
- Figure 1 shows the execution plan for this study

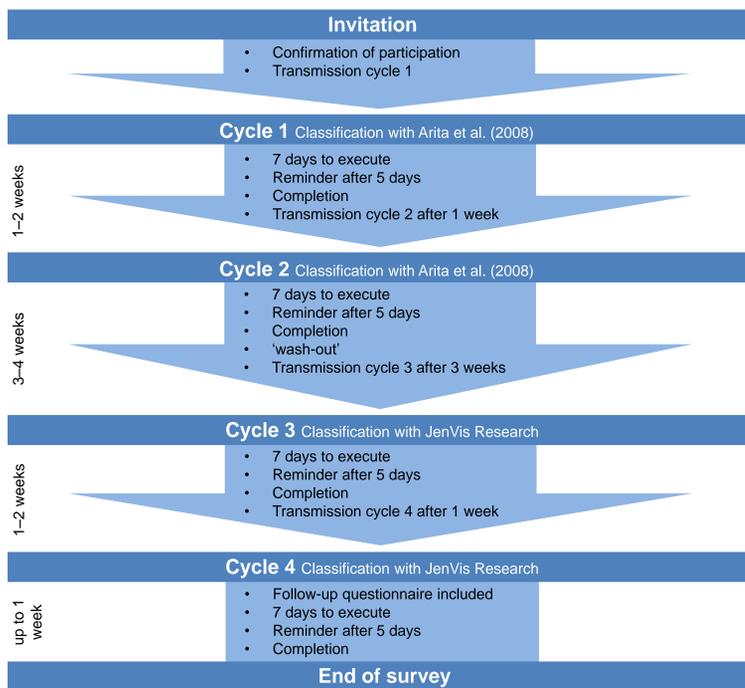


Figure 1: Execution plan for this study

Employed grading scales:

- The well-established grading scale by ARITA et al. [2] (Figure 2, red frame)
- The new illustrative grading scale by JENVIS RESEARCH (Figure 2, blue frame) in dependence on ARITA et al.

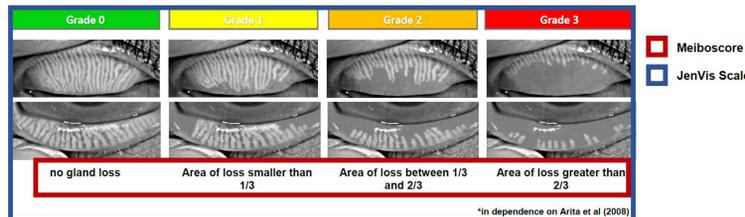


Figure 2: Illustrative JENVIS scale (blue frame) and MEIBOSCORE (red frame) by ARITA et al. (2008) [2]

Statistics → KAPPA-COEFFICIENT according to COHEN and FLEISS

- COHENS-KAPPA measures the level of agreement between repeated classification by one grader and therefore the **intrater reliability** [3]
- FLEISS'-KAPPA measures the level of agreement between classifications of different graders, consequently it measures the **interrater reliability** [3]
- The KAPPA-COEFFICIENTS were interpreted as followed:
KAPPA-Values < 0 (no agreement), 0–0.2 (slight), 0.21–0.4 (fair), 0.41–0.6 (distinct), 0.61–0.8 (substantial), > 0.8 (almost perfect)

Results:

Figure 3 shows the absolute frequency with which the KAPPA values for the intrater reliability were classified according to their level of agreement.

- 48 participants = 48 KAPPA values for both grading scales
- The individual gradings of all 80 images (including upper and lower eyelid) were factored into this calculation of the KAPPA values
- The separate calculations of the KAPPA-COEFFICIENTS for upper and lower eyelid show comparable results

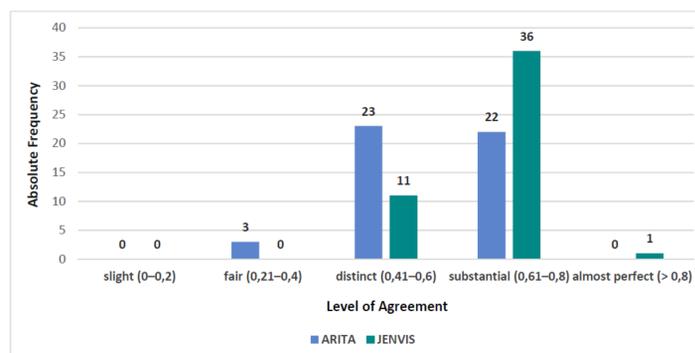


Figure 3: Frequency distribution for the Intrater-KAPPAs (80 images) of the 48 participants, classified according to the level of agreement

Figure 4 shows the box plot diagrams for all six intrater-KAPPA variables.

- The median of the 48 intrater-KAPPAs for the classification with the MEIBOSCORE [2] is 0.57 and therefore shows a 'distinct' agreement between the first and the second evaluation cycle
- Factoring the gradings for the upper and lower eyelid separately into the calculation the agreement is also 'distinct' (upper lid = 0.59; lower lid = 0.56)
- For the classification using the new illustrative grading scale by JENVIS RESEARCH the median of the intrater-KAPPAs is 0.67 which corresponds to a 'substantial' agreement between first and second evaluation cycle
- The same agreement is achieved in the separate calculations for the upper and lower eyelid (upper lid = 0.72; lower lid = 0.63)

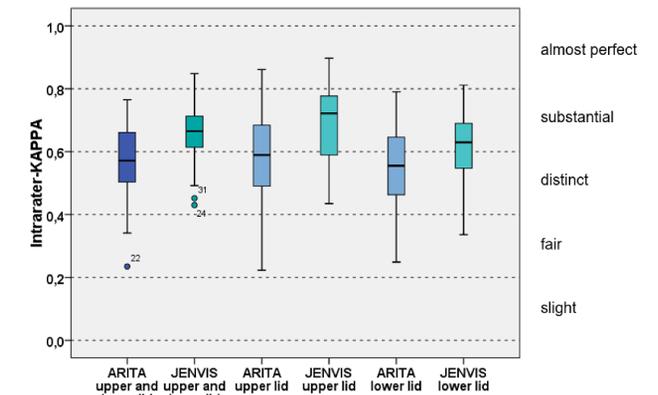


Figure 4: KAPPA-Statistics for the evaluation of the intrater Reliability of the 48 participants in the form of box plots ($\alpha = 0,01$)

- The interrater reliability was calculated with the FLEISS-KAPPA (Table 1)
- It is 0.43 (95 %-CI: 0.28–0.58) for the ARITA scale and 0.52 (95 %-CI: 0.38–0.67) for the JENVIS scale
- Both scales demonstrate a 'distinct' agreement between the classifications of all graders considering all meibography images as well as considering images of the upper and lower eyelid independently

Table 1: Assessment of the FLEISS-KAPPA for the evaluation of the interrater reliability with 95 % confidence interval

| Grading Scale | ARITA Cycle 2 | | JENVIS Cycle 2 | |
|---------------------------|------------------|------------------|------------------|------------------|
| FLEISS-KAPPA with 95 %-CI | 0.43 (0.28–0.58) | | 0.52 (0.38–0.67) | |
| | Upper Lid | Lower Lid | Upper Lid | Lower Lid |
| FLEISS-KAPPA with 95 %-CI | 0.56 (0.4–0.73) | 0.54 (0.38–0.71) | 0.66 (0.5–0.81) | 0.59 (0.43–0.75) |

- The follow-up questionnaire reveals, that the majority of the participants preferred the illustrative grading scale by JENVIS RESEARCH and supports the implementation of additional grading scales for the thickness and tortuosity of the meibomian glands (Figure 5)

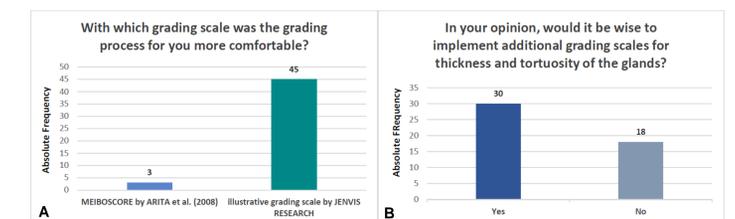


Figure 5: Selected results of the follow-up questionnaire A favoured grading scale B opinion on additional grading scales for tortuosity and thickness of meibomian glands

Conclusion:

Both grading scales show a high degree of agreement concerning the repeatability (intrater reliability) as well as the classification by different graders (interrater reliability).

- Especially the intrater reliability was improved employing the illustrative scale by JENVIS RESEARCH
- Illustrative scales may be beneficial to practitioners to classify meibomian gland dropout.
- The optometric practice is lacking grading scales for the classification of the thickness and tortuosity of meibomian glands

References:

- HEILIGENHAUS, A., J.M. KOCH, D. KEMPER, F.E. KRUSE und T.N. WAUBKE, 1994. Therapie von Benetzungstörungen. Klinische Monatsblätter für Augenheilkunde, 204(3), 162-168
- ARITA, R., K. ITOH, K. INOUE und S. AMANO, 2008. Noncontact infrared meibography to document age-related changes of the meibomian glands in a normal population [online]. Ophthalmology, 115(5), 911-915
- BORTZ, J. und G.A. LIENERT, 2003. Kurzgefasste Statistik für die klinische Forschung. Leitfaden für die verteilungsfreie Analyse kleiner Stichproben