Schimmel Martin (Orcid ID: 0000-0001-9700-5534)
Müller Frauke (Orcid ID: 0000-0003-3981-0134)
Srinivasan Murali (Orcid ID: 0000-0003-3365-576X)
Abou-Ayash Samir (Orcid ID: 0000-0003-1047-5571)

Title page

Title: Assessing masticatory performance with a color mixing-ability test using smartphone camera images.

ID: JOR-22-0078.R1

Running Title: Reliability and comparability of smartphone cameras

Authors: Martin Schimmel[1,2 #]
Elias Rachais[1#]
Nadin Al-Haj Husain[1]
Frauke Müller[2]
Murali Srinivasan[2, 3]
Samir Abou-Ayash[1]

# equal contribution as first author

Affiliations: 1 - Department of Reconstructive Dentistry and Gerodontology, School of Dental Medicine, University of Bern, Bern, Switzerland.
2 - Division of Gerodontology and Removable Prosthodontics, University Clinics of Dental Medicine, University of Geneva, Geneva, Switzerland
3 - Center of General-, Special Care and Geriatric Dentistry, Center of Dental Medicine, University of Zurich, Zurich, Switzerland

Corresponding author:
Martin Schimmel
University of Bern
School of Dental Medicine
Department of Reconstructive Dentistry and Gerodontology
martin.schimmel@zmk.unibe.ch
Phone: +41 31 632 25 86
Fax: +41 31 632 49 33

Data availability statement

The original data of the study can be made available upon- reasonable request. A manual for the Hue-Check Gum © method can be downloaded at

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/joor.13352

https://www.researchgate.net/publication/281447762_Detailed_instructions_for_the_View
Gum_software.

Abstract (250/250)

Background: Color-mixing ability tests are frequently used to assess masticatory performance but the image acquisition process may be cumbersome and technique sensitive.

Objectives: To evaluate the reliability of smartphone camera images in assessing masticatory performance using a color-mixing ability test.

Methods: Participants were recruited into three groups of dental state (n=20 each): fully dentate, removable partial denture wearers and complete denture wearers. After performing a color-mixing ability test, images of the gum specimens (Hue-Check Gum©) were captured with two smartphones and compared with the images obtained from a flat-bed scanner by two examiners. The images were analyzed with a subjective- (SA) and an opto-electronical assessment (VoH). Inter- and intra-rater reliability were tested. ANOVA models with repeated measures were used for statistical analysis (α=0.05).

Results: All three image acquisition techniques were able to distinguish masticatory performance between different dental states. For SA, intra-rater reliability was slight to moderate and inter-rater reliability was substantial to almost perfect. For VoH, intra-rater reliability with the smartphones were significantly different between two examiners, but the inter-rater assessment was reliable. The opto-electronic analysis with smartphone images underestimated the masticatory performance significantly when compared to the flat-bed scanner analysis. Seven-day ageing of the specimens did not significantly affect the results.

Conclusions:

The assessment of masticatory performance with the Hue-Check Gum© is a reliable method. The use of smartphones may occasionally underestimate masticatory performance; image acquisition with a flatbed scanner remains the gold standard. A centralized analysis of the photographed wafer may foster the reliability of the diagnosis.

Keywords: masticatory performance, color-mixing ability, smartphone, Hue-Check Gum, ViewGum

Introduction

Oro-facial fitness, i.e. oro-facial health, was recently conceptualized as “a state that is characterized by an absence of, or positive coping with physical disease, mental disease, pain, and negative environmental and social factors. It will allow natural oro-facial functions such as sensing, tasting, touching, biting, chewing, swallowing, speaking, yawning, kissing, and facial expression”. [1] Especially the chewing function attracts great interest as it may be regarded as a compound outcome parameter of the function of the oro-facial system. In numerous studies it was demonstrated that chewing function is associated with various dental and more general aspects like Oral Health Related Quality of Life [2, 3], ageing [4], cognitive function [5], satisfaction with dental prostheses [6], or the maintenance a healthy diet [7]. Furthermore, it might serve to evaluate dental treatment outcome, e.g. in case of implant-overdentures or removable partial dentures. [8, 9] As chewing function and nutritional status are in part interrelated, impaired mastication may foster the development of frailty and sarcopenia; hence the assessment of masticatory parameters may be important for standard geriatric assessments. [10, 11]

Interest in the evaluation of the chewing function dates back to early days of academic dentistry and was firstly described in 1902. [12] Since then, the methodologies became more and more advanced, but the accepted gold standard for objective evaluation of the chewing function today is still the evaluation of chewing efficiency and chewing performance with breakable test food like peanuts or silicon cubes (for review see [13]). The chewing ability, i.e. the subjective evaluation of chewing by an individual, needs to be differentiated but is an important aspect to fully understand individual chewing behaviors. [14]

There are different types of objectively testing chewing efficiency and/ or chewing performance. However, many require specialized equipment and are cumbersome to perform, like the sieving methods. The two-color mixing test, as originally described by Liedberg and Öwall is based on the ability of an individual to form and knead an elastic two-colored specimen, using wax. [15] Later, chewing-gum was used and based on this principle, Prinz et al. developed an early computer-based analysis to evaluate the degree of color mixture. [16]

Schimmel et al. refined the method in several steps and it is now available with a validated work-flow, comprising of test procedure, image acquisition, software-based evaluation of color mixing ability and a complementary very simple categorical evaluation for use in clinical practice. [17-20]However, this workflow still requires an office/ lab-based flatbed scanner, which may be a hindrance to use in larger epidemiological studies, nursing homes, or in a scientific fieldwork environment. A great facilitator for these tests may be the simplifying of the image acquisition procedures. With the digitalization and development of smartphones using advanced digital cameras, the possibility of reducing the complexity of digitizing the specimens might be drastically reduced.
Hence, the present study aimed to evaluate the use of smartphone cameras in the scope of a previously described two-color mixing ability test to assess masticatory performance. Furthermore, it was aimed to examine the effect of ageing of the specimens. The null hypothesis (H0) was: "There is no difference in the intra- or inter-rater reliability when evaluating two-colored chewing gum specimens using different smartphones compared to the gold standard for image acquisition." Secondary outcome parameters were the discrimination capacity between various dental states and effect of specimen aging in the test outcome.

Material and Methods

The current study protocol was evaluated by the Cantonal Ethics Committee of Bern and it was ruled to not fall under the Swiss Federal Human Research Act (KEK Req-2016-00266) and therefore a formal ethical approval was not required. All study participants were recruited at the School of Dental Medicine, University of Bern, or in a private practice, between January 2019 and October 2020. Eligible subjects were evaluated during routine follow-up appointments, if they fulfilled the general inclusion criteria: age older than 18 years, presented for routine recall appointments and were able to follow the study related instructions.
They were excluded if they presented with orofacial pain, history of severe trauma or had a history of cancer treatment in the orofacial region.

Group-specific inclusion criteria comprised for Group "dentate": fully dentate subjects with a presumed ideal chewing function and therefore presenting with a number of remaining teeth ≥ 28, decayed, missing, filled, teeth index (DMFT) score ≤ 4, and Angle class 1 occlusion. Group

"RPD" were recruited if they wore clinically sufficient bilateral free-end removable partial dentures in at least one jaw. For group "edent", fully edentulous participants with well adapted conventional full dentures were recruited. Age, gender, the number of occluding premolar units (OU), and if present, the type and age of existing dentures was recorded. One OU was counted as a pair of natural premolars in contact in habitual contact position and two OUs for a pair of natural upper and lower molars in occlusion.[21]

As test specimens, two-colored chewing gums with separate blue and pink layers were used (Hue-Check Gum©, University of Bern, Switzerland). This test procedure for assessing masticatory performance itself was validated previously in dentate, partially edentulous and edentulous subjects. [19, 22, 23] The two layers were wetted with water, fused by hand without deforming the two parts, and were placed on the participant's tongue with the blue layer facing downwards. For each participant, one trial of the test was performed to familiarize the subject with the procedure. The participants performed 20 chewing cycles as counted by the operator, and were instructed to chew the specimen as rigorously as possible. Then, the gums were retrieved from the oral cavity, access fluid was removed, and put into a transparent plastic bag. Subsequently, the specimens were flattened to a 1 mm thick wafer, using a resin template. Images of each wafer were obtained from both sides using a flatbed scanner (Epson Perfection V750 Pro; Seiko Epson Corp., Nagano Japan) at a resolution of 300 dots per inch (dpi). Additionally, two generations of mobile phones with disabled flash function (iPhone X; Apple-Corp, Cupertino, CA, USA) at a resolution of 458 pixels per inch (ppi), and an older generation mobile phone (Galaxy A3; Samsung electronics co., Ltd, Suwon, South Korea) were used to obtain images from the two sides specimens. All smartphone pictures were taken free-hand, simulating clinical conditions, at a distance of approximately 10 cm, and ambient light conditions. The wafers were re-digitized and re-photographed within 24 hours after assessment. The chewing gums of the dentate group were stored additionally one week in a standard refrigerator, and scans were obtained again to to evaluate the ageing effects and color stability over time.

Subjective assessment (SA)

The pictures of the specimens were subjected to a subjective assessment using the previously validated ordinal grades[17]: SA 1 chewing gum not mixed, impressions of cusps or folded once, SA 2 large parts of chewing gum unmixed, SA 3 bolus slightly mixed, but bits of unmixed original color, SA 4 bolus well mixed, but color not uniform, SA 5 bolus perfectly mixed with uniform color. [17]

Two independent operators evaluated the images of the wafers captured by the smartphones and the flatbed scanner, to evaluate the inter-rater reliability for SA. One operator repeated the evaluation one week later for the intra-rater reliability. The images of the aged samples were again assessed by both operators to test the effect of ageing on the rating. The two operators were calibrated by the corresponding authors in a pilot experiment using 20 specimens.

Opto-electronic assessment: Variance of Hue (VoH)

The software ViewGum© (www.dhal.com, Athens, Greece) was used for the opto-electronic assessment. [18] The software converts the images of the specimens into the HSI (Hue, Saturation, Intensity) color space and calculates the homogeneity of the color mixture as the Variance of Hue (VoH, range 0 to 1). Well-chewed specimens with a high degree of color mixture present with a low VoH and vice versa. There is a quasi-logarithmic association of VoH and the number of chewing cycles, and masticatory performance. [18, 19] All images of the wafers, mobile phone and scanner, were examined in this way by two operators. One operator evaluated the images twice, once within 24h and the dentate group again one week after storage. The second operator evaluated the images only within the first 24 hours.

Statistical analysis

Sample size estimation was based on previous validation studies with n=20 participants/ group in which the scanning method were able to discriminate between different dental states [17, 19]. Based on the results of these studies a similar sample size of 20 per group was adopted for the current study. Descriptive analyses were performed separately for the overall, group-, assessment-, and operator-wise VoH and the SA values, using means and standard deviations (SDs), medians [quantile25-quantile75], range (minimum, maximum) as appropriate. Weighted kappa was calculated to evaluate inter- and intra-rater agreement and were interpreted according to Landis et al. (1977): Kappa < 0 poor

0-0.2 slight, 0.21-0.4 fair, 0.41-0.6 moderate, 0.61-0.8 substantial, 0.81-1 almost perfect. [24]

A repeated-measures analysis of variance with repeated measures (ANOVA) was used to identify the differences across the groups, the types of assessment, and their interaction.

Subsequently, estimated mean group- and assessment type-wise differences (EMD) were calculated with linear regression analyses. Bland-Altmann plots were used to illustrate the differences in terms of the assessment method. Intra- (repeated measures) and interrater (different operators) were analyzed by means and SDs, limits of agreement, and paired t-tests. All analyses were performed with an alpha of 0.05 using the IBM SPSS 24.0 software (SPSS Inc).

Results

The study sample included sixty volunteers (mean age 57.1 years with a range from 22 to 88 years) with an equal distribution of n=20 per groups dentate, RPD and edent (Table 1).

Subjective assessment (SA)

Overall, the subjective assessment revealed significant differences for the masticatory performance between all groups (all $p<0.001$), the interaction was not significant ($p = 0.998$), indicating that this result does not depend on the assessment method scanner, iPhone or Samsung. The group dentate had the highest masticatory performance, followed by the group RPD and the group edent (Table 2).

The intra-rater reliability was slight to moderate in the groups with the highest degree of color-mixture, i.e. masticatory performance, (group dentate) as assessed on the smart phone camera images and after the ageing procedure. The inter-rater reliability was substantial to almost perfect (Table 3). Nevertheless, great part of the agreement measures in Table 3 were $p>0.05$, which means that there is not enough evidence to conclude that the appraiser agreement is different from what would be achieved by chance.

Opto-electronic assessment (VoH)

The opto-electronic assessment also revealed highly significant differences for the masticatory performance between all groups ($p<0.001$), the interaction was not significant ($p = 0.088$), indicating that this result does not depend on the assessment method scanner, iPhone or Samsung. The group dentate showed the highest masticatory performance, followed by the RPD and the edent groups (Table 4).

For the inter-rater reliability, both in the case of Samsung and iPhone assessment, there are significant differences between the two examiners (Table 5), however the results mostly lie within the limits of agreement (± 2 SD, Figs 1a, 1b). Seven-day ageing of the specimens did not have a significant effect on the results.

When comparing the methods, iPhone and Samsung to the gold-standard scanner, there were significant differences between iPhone and Scanner and between Samsung and Scanner, obtaining results that indicate lower masticatory performance (higher VoH) in the groups dentate and edentate. In the RPD group, only Samsung and Scanner differed significantly. If the pooled data of the three groups were evaluated, the mean results of the iPhone method differed significantly from Scanner and likewise the mean value of Samsung from Scanner. On an average, significantly larger values are obtained with iPhone and Samsung scanner than with the gold-standard (scanner). The upper limit of the 95% confidence interval (VoH) was <0.1 for all comparisons, with smaller upper limits for the iPhone method (Tab. 6).

Discussion

The present study evaluated the reliability and comparability of images of specimens for assessing masticatory performance with a two-color mixing ability test, as recorded with two different mobile phones, and compared to the gold standard, i.e. image acquisition with a flatbed scanner. The subjective assessment and a semi-automatic opto-electronic assessment revealed their feasibility to reliably distinguish between different dental states. However, for the opto-electronic assessment, the analysis with the smart-phone images underestimated in the groups with good chewing function the masticatory performance. Hence, the use of smart-phone images cannot be recommended unconditionally.

For all types of assessment, inter-rater reliability was lower than intra-rater reliability. Although within the limits of agreement of the procedures, in the case of the opto-electronic analysis, there were even significant differences between the smartphones and raters in regard to the assessment of masticatory performance. An ageing period of seven days did not have an influence on the analysis. Therefore, the null-hypothesis can only partly be rejected.

The participants of the current study were divided in three groups with a wide age range and with different intraoral conditions and were recruited from a convenience sample at a University Clinic and private practice. Hence, a good cross-section of possible states of masticatory performance in the sample can be expected. [25, 26] Only few inclusion criteria were defined for the participants. Hence, it can be expected that the findings are generalizable in regard to applied smartphone camera, and individuals with similar dental state. Therefore, the results may be representative for a broader population. However, due the small number of included overall participants (n=60; each group n=20), the findings should not be over-interpreted. Moreover, there is an increased risk of type II error, that is more likely to occur when sample sizes are small, and the true difference or effect between groups and interactions are small and variability is large.

The clinical execution of the mobile phones could have been more standardized, using standard light conditions and standard distances for image recording. [19] Furthermore, evaluating the images by a higher number of operators would have been a useful addition, analyzing the inter- and intra-rater reliability. Stricter eligibility criteria, such as the number of occlusal units, denture age, salivary flow, age, gender or dental status in the antagonizing jaw would have helped to standardize the test procedure even more. However, it should be noted that the main objective of this study was to compare the performance of a flatbed scanner in the analysis of the mixing ability test to that of smartphones, in order to make this analysis practicable also outside a clinical environment. Another weakness is the short-lived production cycle of any digital device and smartphones are a good example for this. Therefore, any research performed with a recent device is deemed to show only a snapshot of the technical development.

Mobile phones with new applications are more and more used as supplement for examinations and research. [27] For assessing chewing function in a clinical environment like nursing homes, or in field studies it can be very helpful to quickly and easily obtain images of the specimens as it if often cumbersome to evaluate the specimens on the spot or transport them to the next available flatbed scanner. Furthermore, a storage of the images might be helpful to follow-up an intervention that might affect the oro-facial function. Therefore, including popular smartphone brands such as the iPhone and an Android Samsung smartphone as examination tools was decided on market share of these devices, which reached 79% in the US in 2021. [28] In addition, for a better discrepancy it was decided to take a new and an old generation mobile phone to cover a bigger area of camera technology for comparing. Given available material, the Samsung was chosen as the older model.

Previous studies showed good agreement between a standard flatbed scanner and mobile phone. [19, 29] However these studies used standardized positions of capturing the wafers. Whereas this study didn't use any standardization except the average position of approximately 10cm distance and normal daily light conditions without considering the time. Furthermore, no special tool was used to edit the pictures taken by the phones. Pictures were only edited with cropping tools using an already installed preview to fit in the ViewGum© software. These decisions were made to represent a real clinical examination without expensive equipment and short evaluation time.

For the Subjective Assessment, the kappa value was moderate to substantial according the comparison of the two operators showing a certain similarity of evaluation. However, subjective components of the operators led to a different observation and lower kappa scores, especially in the groups with removable dentures compared to the dentate group. This leads to the assumption, that the evaluation of the dentate group was easier due to the higher color-mixing of the chewing gum and thus easier score allocation. iPhone and Samsung images were more unreliable to score compared to the scanner image. The standardized resolution of the scanner might have allowed more precise assignment of the images than those of the mobile phones, especially at higher degrees of color mixing. Nevertheless, results showed higher mixed chewed gum was rated by all devices more similar than lower mixed which can be attributed to the subjective component influencing the SA classification. This becomes especially clear when comparing both operators. Individually, higher mixed chewed gums were evaluated with higher SA scores, whereas in comparison less mixed chewed gums were evaluated more similar. Therefore, the results of this study are similar to the findings of the study from. Fankhauser et al. (2020) [29] Pictures from mobile phones were scored not very reliably, and clinicians should be aware that deviations are possible. In addition, as reported by Silva et al. (2018) in a study with complete denture wearers factors such as prevalence bias (the proportion of cases on which the raters agree) may influence the magnitude of the kappa, as well as the poor definition of cut-off criteria for classification of the specimens in the visual analysis. [30]

In regard to the opto-electronical assessment, the method with the Android Samsung smartphone showed throughout the greatest spread and deviation of results. The operators also showed significant differences in the evaluation of the dentate group using the iPhone. It is possible that the non-standardized recording by the phones changed the image quality e.g. due to blurring, camera moving or missed focus. With the scanner, images could always be taken with same lightning and quality and similar findings were also made in a previous study. [29] However, digitizing wafers using a flatbed scanner is time-consuming, and therefore, hardly conceivable in nursing home routines. For practical reasons, it would make sense not to take any photographs at all and to perform the subjective assessment directly based on the chewing gum. However, the present results showed a significant inter-rater variation in the subjective assessment. Therefore, taking photos with a high-quality smartphone camera with subsequent centralized evaluation by a single, calibrated person could be a practical alternative to the scanner method in order to obtain standardized results.

Electronical assessment with a flatbed scanner achieved more reproducible values than with mobile phones, but even more importantly, both smartphones produced images that underestimated the masticatory performance as compared to the gold standard. As the upper limit of the 95% Cis for all comparisons was approximately 0.1, it should be assumed that this possible error must be accepted when using the smartphone cameras. In the analysis of masticatory performance with the Hue-Check Gum© and the ViewGum© software, an error of 0.1 is not negligible and might comprise an error of more than 10%, as VoH ranges from 0 to 1, with most readings lying between 0.05 and 0.8. Imamura et al. found that the cut-off value for oro-facial hypofunction as assessed with the current test may be 0.415, this error could have important effects on the diagnosis of individual patients. [31] However, the underestimation of masticatory performance in a context of long-term care might spark early screening of oro-facial disease by a professional dental care provider and initiate early action to prelude consequences of poor oral health.

Especially in the RPD group, both mobile phone based analyses exceeded the limit of 0.415, whereas the scanner did not. Therefor both mobile phones in this study imply an oro-facial hypofunction, whereas the gold standard analysis would not find such a condition.

This suggests that with the mobile phones, patients are more likely to be classified in the oro-facial hypofunction category, which may affect the course of treatment. Patients who are misclassified may experience overtreatment. Therefore, it is important not to rely only on one evaluation, but to consider other aspects such as further clinical examinations or questionnaires to evaluate the subjective treatment need or alteration of food choice. On the other side, in a clinical environment it could be speculated, that mobile phones evaluation could be used as a quick examination tool for rough estimation to supplement the clinical findings. Additionally, it could be used as a screening for the chewing function.

Unanswered questions and future research

Any smartphone relies on built-in color modification features, and more recently even automated software super-impositions of several pictures (image bursts) to achieve sharp and bright pictures. Hence, smartphone images are always the result of image modifications and do not present a realistic picture as known from analogue photography or an image from a flatbed scanner as used in the current study. [32] It remains nebulous for the everyday user, which image modifications are implemented by the manufacturer and what effect even software updates might have on the image processing. Hence, even with more advanced smartphones in the future, care must be taken to not overestimate their capability in this specific test that relies on the analysis of color distributions. However, for the quick subjective assessment SA, the images as obtained with smartphone images may be sufficient, as there is still a human evaluator to estimate the SA score.

Furthermore, it might be of interest to re-produce the study in individuals with special needs, like dementia patients or children in long-term care facilities as these populations might benefit extensively from simple and practical diagnosis in relation to the oro-facial function. However, the feasibility of these diagnostic tools in special patients groups should be assessed before recommending their use.

Conclusion

The assessment of masticatory performance with the Hue-Check Gum© is a reliable method. Smartphone cameras are readily available and sufficiently precise for clinical use, but may occasionally incite overtreatment by underestimating the masticatory performance. This might however spark early screening of oral disease in frail individuals. A centralized analysis of the photographed wafer may foster the reliability of the diagnosis. Nevertheless, image acquisition with a flatbed scanner remains the gold standard.

Conflict of interest statement

Martin Schimmel is a member of the oral function scientific advisory board for Sunstar Suisse SA (Etoy, Switzerland). All other authors declare no conflict of interest.

Author contributions

Samir Abou-Ayash conceptualized the study protocol and the statistical analysis and contributed significantly to the manuscript. Elias Rachais collected the data and wrote a first draft of the manuscript. Martin Schimmel wrote the concept of the study and finalized the manuscript. Nadin Al-Haj Husain collected data and reviewed the manuscript. Murali Srinivasan provided methodological advice and critically reviewed the final version of the manuscript. Frauke Müller contributed significantly to the manuscript and revised the final version

Figure legends

Figure 1 a, b

Bland-Altman plot for assessing the limits of agreement between the method Samsung, and iPhone respectively, and the gold-standard Scanner. Horizontal line is the mean difference (middle dash line) and the limits of agreement (95% Confidence intervals, upper and lower dash line)

Tables

Table 1.

Basic characteristics of the study sample and separated for the different study groups

Group	Age: mean (min-max)	Gender	Denture age: mean (min-max)	OU: mean (min-max)	Denture location
dentate	23.45 years (22-26 years)	Male: n = 13 Female: n = 7	n.a.	8 (8-8)	n.a.
RPD	71.25 years (51-88 years)	Male: n = 10 Female: n = 10	6.25 years (0-15 years)	0.3 (0-2)	Maxilla: n = 7 Mandible: n = 13
edent	76.5 years (59-88 years)	Male: n = 9 Female: n = 11	3.325 years (0-20 years)	0 (0-0)	Both jaws

dentate = fully dentate, RPD = partially dentate with a removable free-end partial denture in one jaw, edent = fully edentulous, OU = number of occluding premolar units

Table 2:

Difference in masticatory performance in the groups as assessed with the subjective assessment SA

Method	Group	n	SA Median (Q25-Q75)#	Compared to group dentate*	Compared to group RPD *
Scanner	dentate	20	4 (4-4)	-	
	RPD	20	3 (3-4)	-1.1 (-1.9--0.3), p=0.004	
	edent	20	2 (2-3)	-2.6 (-3.6--1.6), p<0.001	-1.5 (-2.3--0.7), p<0.001
Iphone	dentate	20	3 (3-4)	-	
	RPD	20	3 (3-3.5)	-0.8 (-1.5--0.1), p=0.024	
	edents	20	2 (2-3)	-2.3 (-3.1--1.4), p<0.001	-1.4 (-2.2--0.7), p<0.001
Samsung	dentate	20	3 (3-4)	-	
	RPD	20	3 (2.5-3.5)	-0.8 (-1.5--0.1), p=0.025	
	edent	20	2 (2-2.5)	-2.2 (-3.1--1.4), p<0.001	-1.4 (-2.2--0.7), p<0.001
Pooled results**					
	dentate	60	4 (3-4)	-	
	RPD	60	3 (3-4)	-1.0 (-1.4--0.6), p<0.001	
	edentu	60	2 (2-3)	-2.7 (-3.2--2.1), p<0.001	-1.7 (-2.1--1.2), p<0.001

# SA 1=very bad masticatory performance to SA 55= very good masticatory performance
dentate = fully dentate, RPD = partially dentate with a removable free-end partial denture in one jaw, edent = fully edentulous
*ordered probit regression (random effect specimen)
** without aged specimens, adjusted for method

Table 3

Intra- and inter-rater reliability of the Subjective Assessment SA

Group	Method	Equal	Unequal	Match (proportion) observed	Match (proportion) expected	Kappa	p-value
Inter-rater reliability: comparison examiner 1 versus examiner 2 (2nd assessment)							
dentate	Scanner	17	3	0.85	0.48	0.713	0.577
	Iphone	12	8	0.60	0.31	0.422	0.027
	Samsung	10	10	0.50	0.38	0.200	0.062
RPD	Scanner	14	6	0.70	0.37	0.522	0.005
	Iphone	16	4	0.80	0.32	0.705	0.972
	Samsung	13	7	0.65	0.31	0.496	0.741
edent	Scanner	15	5	0.75	0.38	0.598	0.206
	Iphone	14	6	0.70	0.32	0.559	0.124
	Samsung	15	5	0.75	0.39	0.592	0.656
dentate	Scanner aged	13	7	0.65	0.43	0.391	0.079
Intra-rater reliability: comparison 1st and 2nd assessment (examiner 2)							
dentate	Scanner	16	4	0.80	0.49	0.610	0.344
	Iphone	17	3	0.85	0.39	0.755	0.443
	Samsung	19	1	0.95	0.45	0.909	0.310
RPD	Scanner	15	5	0.75	0.34	0.620	0.657
	Iphone	17	3	0.85	0.31	0.783	0.108
	Samsung	17	3	0.85	0.29	0.789	0.109
edent	Scanner	18	2	0.90	0.39	0.836	0.161
	Iphone	20	0	1.00	0.38	1.000	1.000
	Samsung	20	0	1.00	0.41	1.000	
dentate	Scanner aged	17	3	0.85	0.39	0.753	0.084

dentate = fully dentate, RPD = partially dentate with a removable free-end partial denture in one jaw, edent = fully edentulous

Table 4

Results of the opto-electronic assessment as assessed with the Variance of Hue (VoH) parameter

Method	Group	n	VOH mean	SD	Difference to dentate*	Difference to RPD*
Scanner	dentate	20	0.094	0.056	-	-
	RPD	20	0.224	0.168	0.130 (0.049-0.212), p=0.002	-
	edent	20	0.426	0.168	0.332 (0.255-0.409), p<0.001	0.202 (0.094-0.309), p<0.001
Iphone	dentate	20	0.152	0.089	-	
	RPD	20	0.244	0.198	0.093 (-0.008-0.193), p=0.072	
	edent	20	0.491	0.192	0.340 (0.249-0.431), p<0.001	0.247 (0.127-0.367), p<0.001
Samsung	dentate	20	0.148	0.083	-	-
	RPD	20	0.270	0.219	0.123 (0.016-0.230), p=0.025	-
	edent	20	0.497	0.196	0.349 (0.254-0.444), p<0.001	0.227 (0.099-0.354), p=0.001
Pooled results **						
dentate		60	0.131	0.081	-	-
RPD		60	0.246	0.194	0.115 (0.021-0.209), p=0.017	
edent		60	0.471	0.186	0.340 (0.255-0.425), p<0.001	0.225 (0.109-0.341), p<0.001

dentate = fully dentate, RPD = partially dentate with a removable free-end partial denture in one jaw, edent = fully edentulous

* estimated difference with 95% confidence interval and p-value (linear regression with random effect specimen)

** estimated group differences adjusted for method

Table 5

Intra- and inter-rater reliability of the opto-electronic assessment VOH of masticatory performance. Corresponding limits of agreement in Figure 1.

Group	Method	Difference# Mean	SD	95%-CI of Differenz#	p-Value*
Inter-rater reliability: comparison examiner 1 versus examiner 2 (2nd assessment)					
dentate	Scanner	-0.000	0.003	-0.002 ; 0.001	0.709
	Iphone	-0.009	0.028	-0.022 ; 0.004	0.168
	Samsung	-0.008	0.015	-0.015 ; -0.001	0.030
RPD	Scanner	-0.000	0.003	-0.002 ; 0.001	0.549
	Iphone	-0.004	0.008	-0.008 ; -0.001	0.019
	Samsung	-0.009	0.014	-0.016 ; -0.003	0.009
edent	Scanner	-0.000	0.004	-0.002 ; 0.002	0.693
	Iphone	-0.002	0.005	-0.005 ; 0.000	0.073
	Samsung	-0.004	0.012	-0.010 ; 0.001	0.108
dentate	Scanner aged	-0.000	0.005	-0.002 ; 0.002	0.909
Intra-rater reliability: comparison 1st and 2nd assessment (examiner 2)					
dentate	Scanner	0.001	0.004	-0.001 ; 0.003	0.258
	Iphone	0.004	0.025	-0.007 ; 0.016	0.433
	Samsung	-0.000	0.006	-0.003 ; 0.002	0.876
RPD	Scanner	0.001	0.004	-0.001 ; 0.003	0.287
	Iphone	-0.001	0.004	-0.003 ; 0.001	0.171
	Samsung	0.001	0.006	-0.002 ; 0.004	0.378
edent	Scanner	-0.001	0.002	-0.002 ; -0.000	0.050
	Iphone	-0.000	0.003	-0.002 ; 0.001	0.787
	Samsung	0.001	0.004	-0.001 ; 0.003	0.372
dentate	Scanner aged	-0.000	0.004	-0.002 ; 0.002	0.967

dentate = fully dentate, RPD = partially dentate with a removable free-end partial denture in one jaw, edent = fully edentulous

* - t-test for dependent samples

difference in VOH assessments between the two examiners

SD – Standard deviation

Table 6

The opto-electronic assessment VOH in relation to the method of assessment. The scanner method was set as the gold standard. [19, 29]

Group	Method	n	Mean	SD	Difference to Scanner*
dentate	Scanner	20	0.094	0.056	
	Iphone	20	0.152	0.089	0.058 (0.033-0.083), $p<0.001$
	Samsung	20	0.148	0.083	0.054 (0.036-0.073), $p<0.001$
RPD	Scanner	20	0.224	0.168	
	Iphone	20	0.244	0.198	0.020 (-0.009-0.049), $p=0.172$
	Samsung	20	0.270	0.219	0.047 (0.010-0.083), $p=0.012$
edent	Scanner	20	0.426	0.168	
	Iphone	20	0.491	0.192	0.066 (0.036-0.095), $p<0.001$
	Samsung	20	0.497	0.196	0.071 (0.051-0.092), $p<0.001$

* estimated difference with 95% confidence interval and p-value (linear regression with random effect chewing gum), p-value test if the difference is 0

References

[1] Schimmel M, Aarab G, Baad-Hansen L, Lobbezoo F, Svensson P. A Conceptual Model of Orofacial Health with an Emphasis on Function. Journal of oral rehabilitation. 2021;48: 1283-94.

[2] Schimmel M, Leemann B, Christou P, Kiliaridis S, Schnider A, Herrmann F Ret al. Oral health-related quality of life in hospitalised stroke patients. Gerodontology. 2011;28: 3-11.

[3] Stenman U, Ahlqwist M, Bjorkelund C, Hakeberg M. Oral health-related quality of life - associations with oral health and conditions in Swedish 70-year-old individuals. Gerodontology. 2012;29: e440-6.

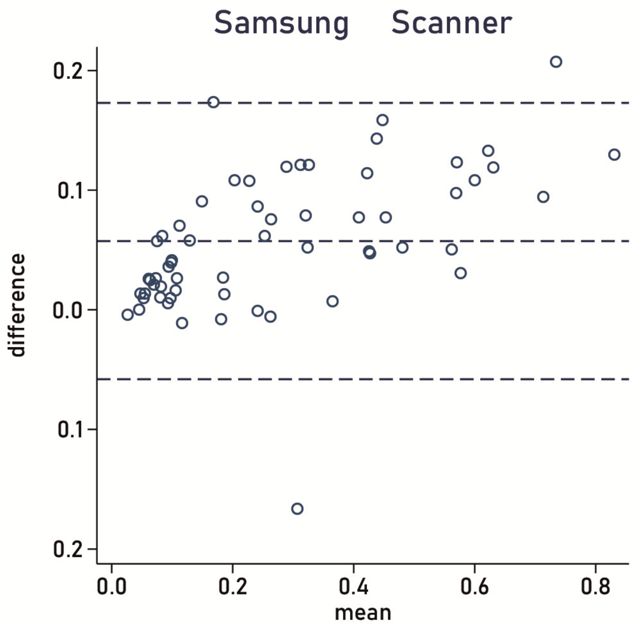
[4] Kosaka T, Ono T, Kida M, Fushida S, Nokubi T, Kokubo Yet al. A prediction model of masticatory performance change in 50- to 70-year-old Japanese: The Suita study. J Dent. 2021;104: 103535.

[5] Elsig F, Schimmel M, Duvernay E, Giannelli S V, Graf C E, Carlier Set al. Tooth loss, chewing efficiency and cognitive impairment in geriatric patients. Gerodontology. 2015;32: 149-56.

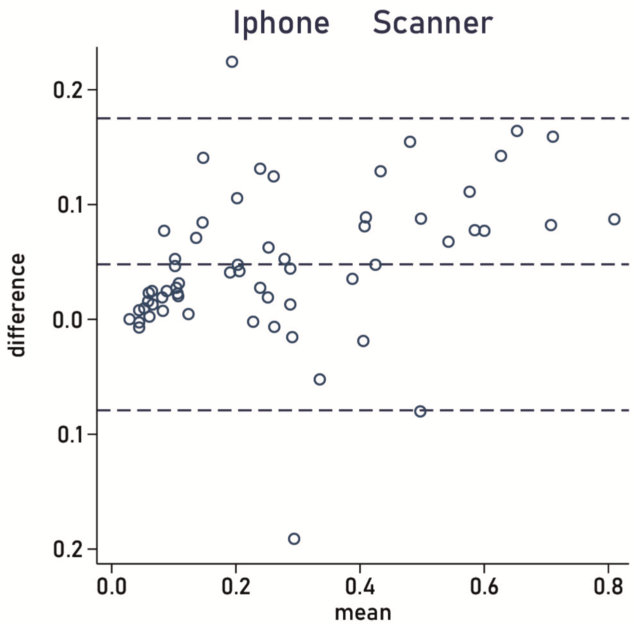
[6] Yoshimoto T, Hasegawa Y, Salazar S, Kikuchi S, Hori K, Ono T. Factors Affecting Masticatory Satisfaction in Patients with Removable Partial Dentures. Int J Environ Res Public Health. 2021;18.

[7] Bousiou A, Konstantopoulou K, Martimianaki G, Peppa E, Trichopoulou A, Polychronopoulou Aet al. Oral factors and adherence to Mediterranean diet in an older Greek population. Aging Clin Exp Res. 2021.
[8] Enkling N, Haueter M, Worni A, Muller F, Leles C R, Schimmel M. A prospective cohort study on survival and success of one-piece mini-implants with associated changes in oral function: Five-year outcomes. Clinical oral implants research. 2019;30: 570-7.
[9] Wallace M, Samietz S, McKenna G, Woodside J V, Schimmel M. Impact of prosthodontic rehabilitation on the masticatory performance of partially dentate older patients: can it predict nutritional state? Results from a RCT. Journal of dentistry. 2018;68: 66-71.
[10] Minakuchi S, Tsuga K, Ikebe K, Ueda T, Tamura F, Nagao Ket al. Oral hypofunction in the older population: Position paper of the Japanese Society of Gerodontology in 2016. Gerodontology. 2018.
[11] Felix Gomez G G, Cho S D, Varghese R, Rajendran D, Eckert G J, Bhamidipalli S Set al. Nutritional Assessment of Denture Wearers Using Matched Electronic Dental-Health Record Data. J Prosthodont. 2022.
[12] Gaudenz J. Ueber die Zerkleinerung und Lösung von Nahrungsmitteln beim Kauact. Arch f Hyg. 1902;39: 230–51.
[13] Goncalves T, Schimmel M, van der Bilt A, Chen J, van der Glas H W, Kohyama Ket al. Consensus on the terminologies and methodologies for masticatory assessment. Journal of oral rehabilitation. 2021;48: 745-61.
[14] Woda A, Hennequin M, Peyron M A. Mastication in humans: finding a rationale. J Oral Rehabil. 2011;38: 781-4.
[15] Liedberg B, Owall B. Oral bolus kneading and shaping measured with chewing gum. Dysphagia. 1995;10: 101-6.
[16] Prinz J, F. Quantitative evaluation of the effect of bolus size and number of chewing strokes on the intra-oral mixing of a two-colour chewing gum. J Oral Rehabil. 1999;26: 243-7.
[17] Schimmel M, Christou P, Herrmann F, Müller F. A two-colour chewing gum test for masticatory efficiency: development of different assessment methods. Journal of oral rehabilitation. 2007;34: 671-8.
[18] Halazonetis D J, Schimmel M, Antonarakis G S, Christou P. Novel software for quantitative evaluation and graphical representation of masticatory efficiency. Journal of oral rehabilitation. 2013;40: 329-35.
[19] Buser R, Ziltener V, Samietz S, Fontolliet M, Nef T, Schimmel M. Validation of a purpose-built chewing gum and smartphone application to evaluate chewing efficiency. Journal of oral rehabilitation. 2018;45: 845-53.
[20] Arakawa I, Abou-Ayash S, Genton L, Tsuga K, Leles C R, Schimmel M. Reliability and comparability of methods for assessing oral function: chewing, tongue pressure, and lip force. Journal of oral rehabilitation. 2020;47: 862-71.
[21] Schimmel M, Leemann B, Herrmann F R, Kiliaridis S, Schnider A, Müller F. Masticatory function and bite force in stroke patients. Journal of dental research. 2011;90: 230-4.
[22] Kaya M S, Guclu B, Schimmel M, Akyuz S. Two-color chewing gum mixing ability test for evaluating masticatory performance in children with mixed dentition: Validity and reliability study. Journal of oral rehabilitation. 2017;44: 827-34.
[23] Barbe A G, Javadian S, Rott T, Scharfenberg I, Deutscher H C D, Noack M Jet al. Objective masticatory efficiency and subjective quality of masticatory function among patients with periodontal disease. J Clin Periodontol. 2020;47: 1344-53.

[24] Landis J R, Koch G G. The measurement of observer agreement for categorical data. Biometrics. 1977;33: 159-74.
[25] Mishellany-Dutour A, Renaud J, Peyron M A, Rimek F, Woda A. Is the goal of mastication reached in young dentates, aged dentates and aged denture wearers? Br J Nutr. 2008;99: 121-8.
[26] Peyron M A, Blanc O, Lund J P, Woda A. Influence of age on adaptability of human mastication. J Neurophysiol. 2004;92: 773-9.
[27] Dorsey E R, Yvonne Chan Y F, McConnell M V, Shaw S Y, Trister A D, Friend S H. The Use of Smartphones for Health Research. Acad Med. 2017;92: 157-60.
[28] O'Dea S. Manufacturers' market share of smartphone sales in the United States from 1st quarter 2016 to 2nd quarter 2021 https://www.statista.com/statistics/620805/smartphone-sales-market-share-in-the-us-by-vendor/. accessed on 21.12., 2021.
[29] Fankhauser N, Kalberer N, Muller F, Leles C R, Schimmel M, Srinivasan M. Comparison of smartphone-camera and conventional flatbed scanner images for analytical evaluation of chewing function. Journal of oral rehabilitation. 2020;47: 1496-502.
[30] Silva L, Nogueira T, Rios L, Schimmel M, Leles C. Reliabilty of a two-colour chewing gum test to assess masticatory performance in complete denture wearers. Journal of oral rehabilitation. 2018;45: 301-7.
[31] Imamura Y, Ohta M, Chebib N, Schulte-Eickhoff R, Schimmel M, Graf Cet al. Masticatory assessment for scoring Oral Hypofunction: alternative methods. Gerodontology. under revision;GER-21-SI-3211.
[32] Dudhane A, Zamir S, Khan S, Khan F, Yang M. Burst Image Restoration and Enhancement. Computer Vision and Pattern Recognition. 2021;arXiv:2110.03680.



JOOR_13352_Fig 1 a VOH Samsung_Scanner.jpg



JOOR_13352_Fig 1 b VOH iPhone_Scanner.jpg