applied sciences

MDPI

Article

Whitening Toothpaste Alone or with a Home-Use or Professional Polishing Paste on a Stained Composite In Vitro

Merve Şahin [1], Ece Altınışık [1], Asmat Asya Ali [1], Ezgi Erden Kayalıdere [1] and Zeynep Hale Keles [2,*]

[1] Department of Restorative Dentistry, Faculty of Dentistry, İstanbul Kent University, 34433 İstanbul, Türkiye; 210301204@mezun.kent.edu.tr (E.A.); 2451610004@ogr.kent.edu.tr (A.A.A.); ezgi.kayalidere@kent.edu.tr (E.E.K.)

[2] Department of Restorative Dentistry, Faculty of Dentistry, İstanbul Atlas University, 34408 İstanbul, Türkiye

* Correspondence: hale.keles@atlas.edu.tr; Tel.: +90-850-450-34-39

Featured Application

These findings describe how a whitening toothpaste, used alone or combined with a home-use or a professional polishing paste, changed the colour, whiteness and surface roughness of one stained nanohybrid composite in vitro. They are intended to inform, not to prescribe, the choice of such regimens; clinical recommendations would require confirmation in designs that measure the relevant clinical endpoints.

Abstract

Whitening toothpastes are widely used by patients with composite restorations, and polishing pastes are increasingly applied to the same surfaces. Forty nanohybrid composite discs (8 × 2 mm) were polished, allocated to four groups (n = 10) and stained in coffee for 144 h at 37 °C. Brushing representing an assumed one month of use was performed with a powered toothbrush (4 min); combined protocols added a second product (home-use paste 64 s, prophylaxis paste 12 s). Surface roughness (Ra) and CIELAB coordinates were recorded at baseline (T0), after staining (T1) and after brushing (T2). Because the intervention occurred only between T1 and T2, ΔRa was the primary roughness outcome; colour recovery (ΔE00, T0 → T2) and net whiteness change (ΔWID) were the primary colour outcomes. On ΔRa, the prophylaxis paste alone roughened the surface relative to the control (p = 0.005) and alone left specimens whiter than baseline; its mean post-brushing Ra reached the 0.2 µm plaque-retention threshold, the confidence interval including that value. Adding the home-use paste changed neither roughening nor net whiteness but left a larger residual colour difference than the toothpaste alone. Whiteness gain and residual mismatch favoured different arms, and the arm that whitened also roughened.

Keywords: composite resins; dentifrices; dental prophylaxis; surface properties; colour; toothbrushing; whiteness index

Check for updates

Academic Editor: Amaury Pozos-Guillen

Received: 8 September 2026
Revised: 6 October 2026
Accepted: 11 October 2026
Published: 13 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Whitening toothpastes are the most widely used response to the growing demand for a brighter smile, acting mainly through the mechanical cleaning power of abrasive particles and, to a lesser extent, through chemical agents [1,2]. Abrasive content cuts both ways: while removing extrinsic stains, it also affects restorative surfaces. Because composite resins consist of hard fillers in a softer organic matrix, the matrix may be worn selectively, and

protruding fillers may increase roughness. Bollen et al. [3] defined a threshold surface roughness of 0.2 μm for bacterial plaque retention on intra-oral hard surfaces, below which no further reduction in plaque accumulation is expected and above which plaque accumulation increases.

Professional polishing and prophylaxis pastes are also used on stained surfaces, but their effect on composite resin is inconsistent: some report that the mechanical polishing action reduces roughness [4,5], others a distinct increase [6,7]. The inconsistency has been attributed to the abrasive system rather than to its declared abrasivity alone: the size, shape and hardness of the particles, and the way they meet a surface made of hard filler in a softer matrix, govern whether material is removed evenly or the matrix is worn preferentially [8]. Liebermann et al. [4] showed by electron microscopy that round, small particles leave smoother surfaces, and Kurucu Karadeniz et al. [9] found a toothpaste of lower declared abrasivity to roughen more than one of higher abrasivity. Similar formulations are now also marketed for home use, blurring the boundary between the two product classes.

For objective colour assessment, the CIEDE2000 formula is preferred, reflecting the perceptual characteristics of the human eye better than CIELAB [10,11]; Paravina et al. [12] established 50:50 perceptibility and acceptability thresholds of $\Delta E00 = 0.8$ and 1.8. To express whiteness on a single scale, Pérez et al. [13] developed a whiteness index for dentistry (WID) and later reported its thresholds (WPT = 0.72; WAT = 2.62) [14].

Whitening toothpastes have been examined extensively on composite resins [9,15–17], and one of the products used here as a home-use polishing paste has already been examined in a toothpaste comparison [17]; whiteness indices have begun to be reported alongside CIEDE2000 on stained composites [7,18]. What remains unresolved is whether adding a polishing paste to a whitening toothpaste changes the outcome on a stained composite, and whether the answer is the same for a product used repeatedly at home as for a single short professional application. The two paste regimens differ in class, duration and setting and are therefore not compared with one another here; each is compared with the toothpaste alone and with an untreated control. Reading the residual difference from baseline together with the direction of the whiteness change and with surface roughness allows a regimen that leaves a surface whiter to be distinguished from one that leaves it closer to its original colour—a distinction that a single unsigned colour difference cannot make. The closest precedent is the work of Meseli et al. [17], in which five composites were stained in coffee for 144 h and brushed in a force-guided simulator with toothpastes of graded abrasivity, the same home-use polishing paste being added once a week. The present design differs in using a powered toothbrush rather than a simulator, in applying the polishing paste twice weekly, and in adding a single short application of a professional prophylaxis paste as a separate arm. This study applied that reading to a coffee-stained nanohybrid composite. The null hypothesis was that the four groups would not differ.

2. Materials and Methods

The design is summarised in Figure 1 and proceeded in five stages. Forty discs were made from one composite lot and finished and polished by one procedure; they were allocated to four groups of ten; baseline surface roughness and colour were recorded (T0); all forty were stained in the same coffee solution for 144 h and measured again (T1); each group then received its assigned treatment—distilled water, the whitening toothpaste alone or the toothpaste followed by one of the two polishing pastes—and was measured a third time (T2). The treatments therefore differ only between T1 and T2; every specimen followed one identical protocol up to T1.

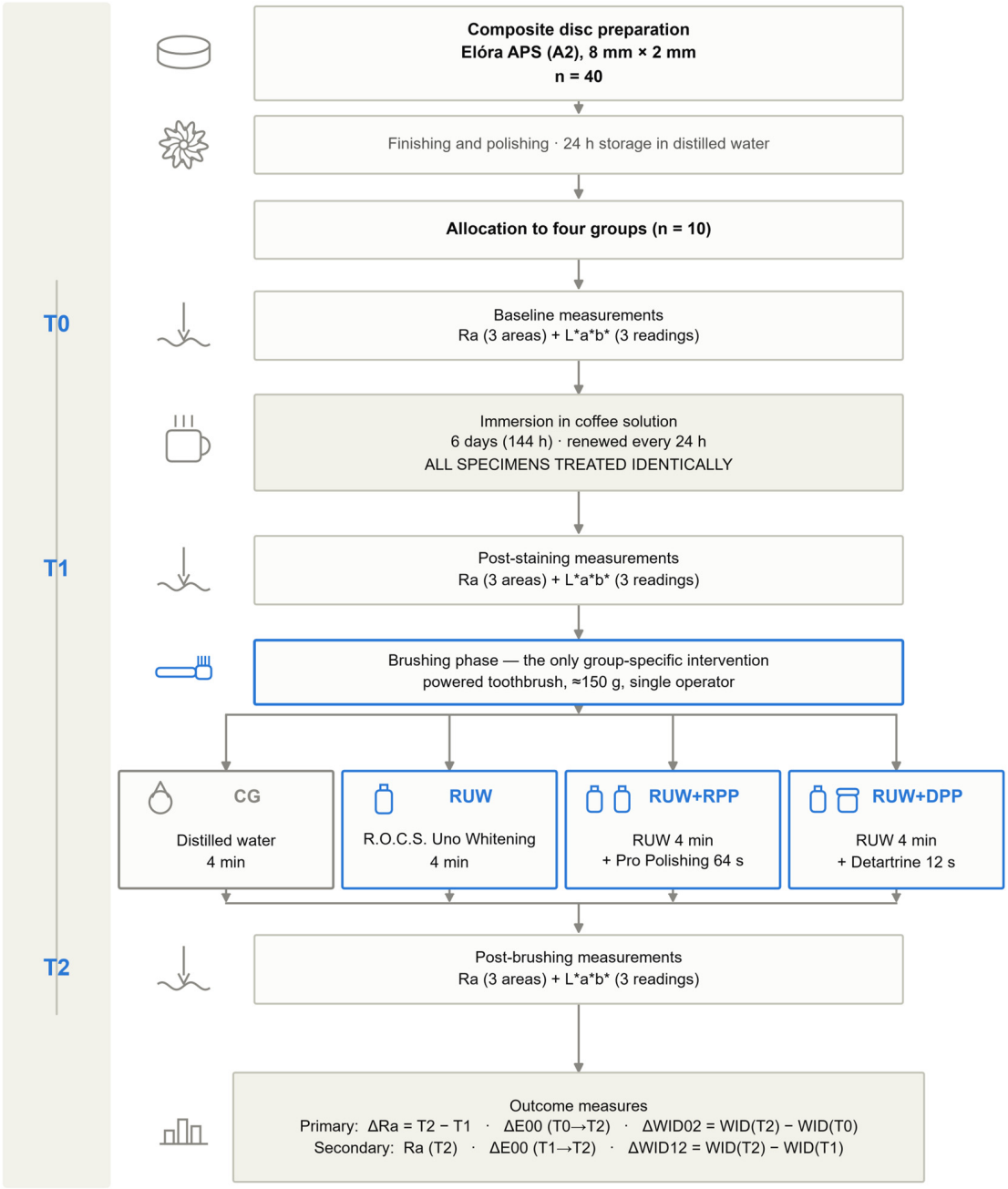


Figure 1. Flow chart of the study. CG, distilled-water control; RUW, whitening toothpaste alone; RUW + RPP, toothpaste plus the home-use polishing paste; RUW + DPP, toothpaste plus the professional prophylaxis paste. T0, baseline; T1, after staining; T2, after brushing. Ra, arithmetic mean surface roughness; L*, a*, b*, the CIELAB lightness, red–green and yellow–blue coordinates. The group-specific intervention occurred only between T1 and T2; up to T1, all specimens underwent identical procedures, which determined the primary outcomes. Allocation preceded baseline measurement.

2.1. Sample Size

Sample size was determined a priori with G*Power 3.1.9.7 (Universität Düsseldorf, Düsseldorf, Germany) for a one-way ANOVA [19] with four groups. The calculation was intended to support the omnibus comparison of the four regimens on surface roughness, the outcome for which effect sizes were available in the literature. No pilot data of our own were available, so a large group effect was assumed (Cohen $f = 0.55$; $\eta^2 = 0.23$); this assumption is compatible with the pairwise roughness differences reported by Kurucu Karadeniz et al. [9] and Gömleksiz and Okumuş [15], which correspond to $d = 3.88$ and

d = 1.90, respectively, although those are pairwise rather than omnibus effect sizes. With α = 0.05 and 80% power, the calculation returns 41 as the smallest total sample reaching that power (0.810); the 40 specimens prepared (10 per group) give 0.798, which is 0.80 to two decimals. This one-specimen difference is not the limiting feature of the design. The more relevant limitation is the precision available for the clinically meaningful pairwise contrasts, in particular the incremental effect of the home-use paste: for a medium pairwise difference (d = 0.8), power is only 40%. Confidence intervals are therefore reported for all primary pairwise contrasts, and non-significant comparisons are described as differences not detected rather than as equivalence, for which no margin was specified. Group size is consistent with comparable studies (6–12 per group) [1,5,16–18]. Post hoc power is a monotone function of the p value and was not reported.

2.2. Specimen Preparation

A nanohybrid composite resin in shade A2 with a declared filler load of 76 wt% and a declared particle size distribution of 0.02–0.7 µm (Elóra APS, body; FGM Dental Group, Joinville, SC, Brazil) was used; both figures are the manufacturer's data and were not verified here. Custom silicone moulds 8 mm in diameter and 2 mm deep were used for standardisation. A microscope slide and a transparent matrix strip (Mylar strip; SS White Dental, Lakewood, NJ, USA) were placed at the base of the mould, the material was inserted and the top surface covered with a transparent strip, pressed with a glass slide to extrude excess material and prevent an oxygen-inhibition layer. Polymerisation was performed from both surfaces for 20 s each with an LED unit (Bluephase PowerCure; Ivoclar Vivadent AG, Schaan, Liechtenstein) emitting at 390–500 nm and 1200 mW/cm^2, with the tip perpendicular and in contact. Irradiance was verified with the radiometer integrated in the charging base.

Specimens were stored in distilled water for 24 h for post-polymerisation stabilisation, then finished and polished with spiral rubber polishers (EVE Diapol; EVE Ernst Vetter GmbH, Keltern, Germany) at the manufacturer-recommended 7000–12,000 rpm, dry, for 10 s each; polishing time was the same for every specimen, but the applied load and the speed used within that range were not recorded, although both affect roughness and gloss [20]. The 40 specimens were allocated to four groups (n = 10) and kept in distilled water at room temperature except during measurement and application. Allocation was intended to be random, but the method used was not recorded at the time, and no allocation list was retained; the procedure is therefore not reconstructed here. What the contemporaneous record does establish is the following. All specimens were made from a single composite lot (050825), so no material batch can coincide with group assignment. Every specimen underwent an identical protocol from fabrication to the T1 measurement, the group-specific product being introduced only afterwards. The coffee solution was prepared to one recipe and renewed on the same 24 h schedule for all specimens, and each specimen was immersed in its own well of a multi-well plate rather than in a vessel shared with other specimens, so a shared staining container cannot account for a group-wide pattern. Brushing was performed throughout by a single operator. Specimen identity was preserved across the three time points, so the paired T0, T1 and T2 readings belong to the same disc. The record does not establish the allocation sequence, the order of specimen fabrication and polishing, the position of a specimen within a plate, the volume of solution per well or the order of measurement; no blinding of the operator to group assignment was implemented.

2.3. Groups

The four groups, the products applied and their abrasive characteristics are summarised in Table 1.

Table 1. Study groups and product characteristics.

Group	Application	Product Type	Manufacturer	Abrasive Content	RDA	Paste Time
CG	Distilled water (control)	—	—	—	—	—
RUW	R.O.C.S. Uno Whitening	Home-use whitening toothpaste	WDS, Moscow, Russia	Silica; xylitol, sodium glycerophosphate	— *	—
RUW + RPP	RUW + R.O.C.S. Pro Polishing	Home-use polishing paste	WDS, Moscow, Russia	Silica (2–3 µm), hydroxyapatite	≈105 †	64 s
RUW + DPP	RUW + Detartrine	Professional prophylaxis paste (clinical use)	Septodont, Saint-Maur-des-Fossés, France	Silicon dioxide, glycerin	150 ‡	12 s

Product names, manufacturers, product classes and declared abrasive constituents are taken from the manufacturers' product information and were not verified analytically in this study; the paste application times are the protocol of this study, not manufacturer instructions. RDA values are literature values rather than manufacturer declarations. * No RDA is stated by the manufacturer of R.O.C.S. Uno Whitening, and none was supplied on request, so the abrasivity of the toothpaste common to the three treatment groups is not known, and the four protocols cannot be ranked by abrasivity. † Approximately 105, the value reported by Meseli et al. [17] for R.O.C.S. Pro Polishing and for the toothpaste-plus-paste regimens containing it; it was not measured in that study either. ‡ 150, the value reported by Mocny-Pachońska et al. [21] for Detartrine. The particle size given for the silica in R.O.C.S. Pro Polishing (2–3 µm) is likewise taken from Meseli et al. [17]. Lot numbers: composite 050825; RUW L530; RPP L0376; DPP B31296AA. Paste time is in addition to the 4 min toothpaste brushing.

2.4. Staining

Baseline colour values were recorded before staining (T0). Specimens were then immersed in 2 g soluble coffee (Nescafé Classic; Nestlé, Vevey, Switzerland) dissolved in 200 mL distilled water (90 °C) and kept at 37 °C for 144 h, the solution being renewed every 24 h. Specimens were then rinsed, blotted dry and measured (T1). Coffee is the staining agent most frequently used in in vitro colour stability research, although no protocol has been agreed upon for solution temperature, immersion time and renewal interval [22]. Soluble coffee was chosen because it dissolves completely, so the concentration is set by weight and is reproduced exactly at each renewal and in each well; a brew from ground beans carries suspended solids that settle and adhere unevenly, which would add a source of between-specimen variation without making the model more severe. The concentration and the renewal schedule follow the practice of comparable in vitro studies rather than a single reference protocol, none having been agreed upon [22].

2.5. Brushing Protocol

Brushing was performed by a single operator with a powered toothbrush with a pressure sensor (Oral-B Triumph 5000, 3D White head; Braun GmbH, Kronberg, Germany). Before brushing, the force applied by the operator was set against a precision balance: the specimen holder was placed on the balance, and the operator adjusted the applied pressure until the reading corresponded to approximately 150 g (≈1.47 N), within the 0.5–2.5 N range specified by ISO standards for wear studies [20]. This was done once, before the brushing sessions, and was not repeated during them. Toothpastes were diluted 1:1 by weight with distilled water to simulate dilution by saliva [18]. Published estimates of daily contact time on a single tooth surface range from a maximum of 5 s for 2 × 120 s of daily brushing [23] to 10 s applied as 5 s morning and evening [17]; brushing time and load determine surface deterioration [24]. Four minutes of continuous brushing therefore corresponds to roughly three to seven weeks of use, and approximately one month was

adopted as a working assumption about cumulative exposure, not a validated reproduction of one month of clinical brushing behaviour. In the combined protocols, the second product was applied separately: 64 s for the home-use polishing paste and 12 s for the prophylaxis paste, the 64 s corresponding to twice-weekly use over the simulated month, the upper end of the manufacturer's recommended frequency. Specimens were then rinsed, dried and measured (T2).

2.6. Surface Roughness Measurement

Roughness was measured with a contact profilometer (Mitutoyo Surftest SJ-410; Mitutoyo Corp., Kawasaki, Japan) with a 2 μm (60°) diamond stylus and 0.75 mN measuring force. Sampling conditions followed ISO 4288:1996 [25] for the expected range (Ra 0.1–2 μm): cut-off (λc) 0.8 mm, evaluation length 4 mm, Gaussian filter. ISO 4288:1996 has since been cancelled and replaced by ISO 21920-3:2021 [26]. The instrument was calibrated against its reference specimen before each measurement session. Because roughness may not be homogeneous, three independent areas were measured per specimen and averaged; baseline measurement is required for a sensitive analysis of surface change [27]. Measurements were repeated at T0, T1 and T2.

Because the readings were taken from different areas rather than the same point, their dispersion reflects the spatial heterogeneity of the surface rather than instrument uncertainty. The within-specimen standard deviation was 0.049–0.057 μm across the three time points (a limits-of-agreement type spread of 0.135–0.158 μm), so a single disc carries a relatively wide uncertainty whereas group means are estimated far more narrowly.

2.7. Colour and Whiteness

Colour was measured with a digital spectrophotometer (VITA Easyshade V; VITA Zahnfabrik, Bad Säckingen, Germany), calibrated against its own calibration block before each measurement. Values were computed for the CIE D65 illuminant and the CIE 1931 2° standard observer; the illumination and detection geometry is that of the instrument's contact probe and is not user-selectable. The 5 mm measuring tip was positioned to rest entirely within the 8 mm specimen surface. Consecutive readings at the same position were averaged; the number of readings for each specimen and time point is given in Table S2. Measurements were made under constant ambient lighting with the specimen resting on a neutral grey background. Specimens were blotted dry and measured within 2 min of removal from distilled water, so that the surface was free of visible water without being allowed to dehydrate. Repeatability of this device family has been reported [28], but validation of a device family does not quantify uncertainty under the present conditions, so it was estimated from the repeated readings themselves. Because these were taken consecutively at one position, their dispersion is instrument repeatability rather than surface heterogeneity. The pooled within-specimen standard deviation was 0.46–0.83 for L*, 0.12–0.23 for a* and 0.47–0.67 for b* across the three time points, giving 0.84–1.04 WID units for a single reading. The uncertainty of a single specimen's whiteness is therefore of the same order as the spread of the group means (0.66 versus 1.66 WID units at T2), whereas the corresponding standard error of a group mean of ten specimens is 0.21 WID units; group comparisons are accordingly little affected, individual specimen values much more so. The same distinction applies to the roughness measurements (Section 2.6).

Colour differences were calculated with CIEDE2000 using the corrections defined by Sharma et al. [10], with kL = kC = kH = 1. Three differences were defined: ΔE01 (T0 → T1, amount of staining), ΔE12 (T1 → T2, stain removal) and ΔE02 (T0 → T2, return to baseline colour). ΔE02 was chosen as the primary colour outcome because it expresses the final aesthetic state and, unlike ΔE12, does not contain the stained reading itself; it is nevertheless

not independent of how heavily a specimen stained (r(ΔE01, ΔE02) = +0.34; p = 0.034) and is interpreted accordingly. Perceptibility (PT = 0.8) and acceptability (AT = 1.8) thresholds were applied [12]. Whiteness was calculated as WID = 0.511·L* − 2.324·a* − 1.100·b* [13]. Net change from baseline (ΔWID02 = WID(T2) − WID(T0)) and gain relative to the stained state (ΔWID12 = WID(T2) − WID(T1)) were defined, with WPT = 0.72 and WAT = 2.62 as thresholds [14].

2.8. Statistical Analysis

Outcome selection followed the time structure of the study: the group-specific intervention occurred only between T1 and T2, whereas all specimens underwent identical procedures from T0 to T1. The primary roughness outcome was therefore the change during that phase (ΔRa = T2 − T1); the post-brushing level (T2) was secondary, since it also carries differences present on entering that phase. For the same reason, the primary whitening outcome was net change from baseline, with baseline whiteness as covariate; ΔWID12 is coupled to the degree of staining and was reported as secondary only.

Normality was assessed by the Shapiro–Wilk test [29], which is preferred at the present group size of ten [30], and homogeneity of variance by the Levene test; the median-based statistic is reported throughout, with the mean-based value given alongside it where the two differ materially [31]. Groups were compared by one-way ANOVA with Tukey HSD [19]; where variances were not homogeneous, results were confirmed by Welch ANOVA [32] and the Kruskal–Wallis test [33], and pairwise comparisons were made with the Games–Howell procedure [34], which does not assume equal variances or equal group sizes [35]. Effect size was reported as η^2. Because the degree of staining (ΔE01) differed among groups, ΔE12 was analysed by covariance analysis with ΔE01 as covariate and homogeneity of slopes verified. The three time points were examined with a repeated-measures analysis of variance, with group as the between-subjects factor and time as the within-subject factor. Sphericity was not met for either outcome (Mauchly p = 0.012 for roughness and $p < 0.001$ for whiteness); the uncorrected tests are reported, since the Greenhouse–Geisser correction (ε = 0.82 and 0.70) leaves every effect significant at $p < 0.001$. Their residuals were approximately normal for roughness (Shapiro–Wilk p = 0.43) but not for whiteness (p = 0.022), and residual variance differed across the time points; these repeated-measures results are therefore reported as a description of the time course, the primary inference resting on the three outcomes above. The analysis-of-variance tables for the three primary outcomes and the corresponding residual diagnostics are given in Appendix A (Table A1 and Figure A1). Significance was set at α = 0.05.

Because the groups were not balanced before the intervention, the imbalance was addressed in four steps. First, the primary outcomes were defined so as to depend on it as little as possible: the roughness outcome is the change during the phase in which the intervention occurred, the primary colour outcome requires no covariate and the primary whitening outcome is adjusted for baseline whiteness. Second, each was repeated in a sensitivity analysis using the immediately preceding time point as covariate (T2 roughness adjusted for T1; net whiteness change adjusted for WID(T1); stain removal adjusted for ΔE01). Third, adjusted group estimates, all pairwise contrasts and their Tukey-corrected 95% confidence intervals are reported, since a significant omnibus test does not establish that individual pairwise differences persist after adjustment. Fourth, covariate overlap between groups was inspected, and any reliance on extrapolation is reported with the corresponding estimates. Three outcomes were treated as primary—ΔRa, ΔE02 and ΔWID02—and each omnibus test meets a Bonferroni-corrected threshold of 0.05/3; all remaining comparisons, including the imbalance-directed analyses, are reported as exploratory and are not corrected for multiplicity. Analyses were performed with IBM SPSS Statistics 31 (IBM Corp.,

Armonk, NY, USA; proprietary software, https://www.ibm.com/products/spss-statistics, accessed on 8 September 2026). The covariate-adjusted pairwise contrasts, the confidence intervals and the model diagnostics were computed with Python 3.13 and the open-source packages statsmodels 0.14 and SciPy 1.18 (https://www.statsmodels.org, scipy.org, accessed on 8 September 2026); both the SPSS syntax and the Python scripts are provided as Supplementary Materials.

3. Results

3.1. Surface Roughness

At baseline, the groups did not differ significantly ($F(3,36) = 2.27$; $p = 0.097$; Table 2), although this is not evidence of equivalence: the largest pairwise difference was 0.033 µm, which 10 specimens per group cannot exclude. After staining, a clear difference appeared ($p < 0.001$), arising entirely from the RUW + RPP group, and the T0–T1 change also differed ($p = 0.023$). Since all specimens had undergone identical procedures to this point, the difference at T1 cannot be an intervention effect: the groups did not enter the brushing phase from the same starting point.

Table 2. Surface roughness (Ra) at three time points and change during the brushing phase.

Group	*n*	T0 (µm)	T1 (µm)	T2 (µm)	ΔRa (T2 − T1) *	Post Hoc	T2 < 0.2 µm
CG	10	0.114 ± 0.024	0.104 ± 0.019	0.106 ± 0.036	+0.003 ± 0.040	b	10/10
RUW	10	0.115 ± 0.022	0.107 ± 0.014	0.145 ± 0.032	+0.038 ± 0.030	ab	10/10
RUW + RPP	10	0.135 ± 0.040	0.165 ± 0.048	0.189 ± 0.035	+0.024 ± 0.053	b	5/10
RUW + DPP	10	0.102 ± 0.024	0.096 ± 0.030	0.204 ± 0.046	+0.108 ± 0.071	a	4/10
p		0.097	<0.001	<0.001	<0.001		

Mean ± standard deviation. * Primary outcome ($F(3,36) = 8.12$; $p < 0.001$; $\eta^2 = 0.404$); letters denote Games–Howell groupings for ΔRa ($p < 0.05$), used because variances were unequal (Levene test based on the median, $p = 0.016$; based on the mean, $p = 0.010$). The T2 level is a secondary outcome ($F(3,36) = 13.81$; $p < 0.001$), and the two measures rank the middle groups differently (see text). The final column counts specimens strictly below 0.2 µm; one further RUW + RPP specimen lay exactly at 0.200 µm.

The primary outcome, roughness change during the brushing phase (Figure 2), differed significantly among groups ($F(3,36) = 8.12$; $p < 0.001$; $\eta^2 = 0.404$). Normality was met, but variances were not homogeneous (Levene $p = 0.016$); tests not assuming equal variances agreed (Welch $F(3, 19.2) = 5.52$; $p = 0.007$; Kruskal–Wallis $p = 0.006$), and pairwise comparisons used Games–Howell. RUW + DPP differed from the control ($p = 0.005$) and from RUW + RPP ($p = 0.036$), while its difference from RUW alone was borderline ($p = 0.058$); no other pair differed ($p \geq 0.148$). Neither the toothpaste alone nor the added home-use paste differed from the control, so on this phase-specific outcome, the prophylaxis paste was the only application producing measurable roughening relative to the control. The adjusted analysis reported below ranks the two paste arms differently, and both are reported. The group × time interaction was significant ($F(6,72) = 7.95$; $p < 0.001$).

As a secondary outcome, the post-brushing level also differed ($F(3,36) = 13.81$; $p < 0.001$). RUW + DPP was highest and CG lowest on both measures, but the two middle groups ranked differently because the RUW + RPP specimens were already rougher before brushing. All CG and RUW specimens remained below the 0.2 µm threshold, against 5/10 in RUW + RPP—one further specimen lay exactly at 0.200 µm—and 4/10 in RUW + DPP. The RUW + DPP mean (0.204 µm) reached the threshold but its 95% confidence interval (0.171–0.237 µm) includes it, so it cannot be said to exceed it; with three of four arms below, clinical relevance should be read with restraint.

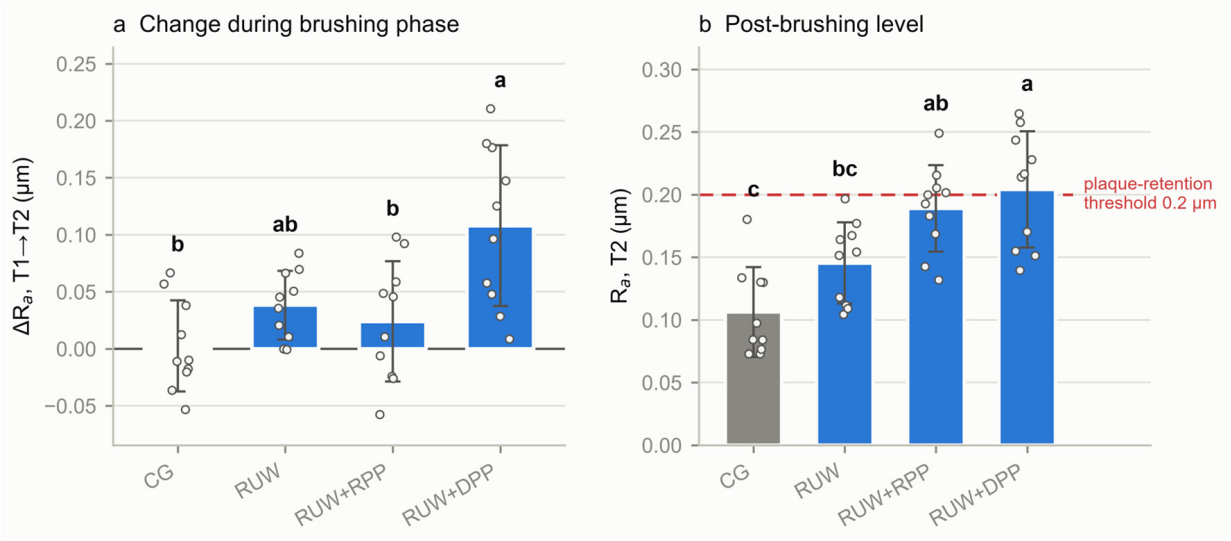


Figure 2. Surface roughness. (**a**) Primary outcome: change during the brushing phase (ΔRa = T2 − T1); the horizontal line marks no change. (**b**) Secondary outcome: post-brushing level; the dashed line is the 0.2 µm threshold. Bars are means, whiskers standard deviations, circles individual specimens; different letters denote significance ($p < 0.05$; Games–Howell in (**a**), Tukey HSD in (**b**)).

Sensitivity analyses did not alter the conclusions: using the median, lowest or highest reading instead of the mean, and using change from baseline, the group effect remained significant with the same ranking. No specimen was flagged as an outlier, and none was excluded. Because absolute values shift with the summary method (0.196–0.258 µm for RUW + DPP), the 0.2 µm threshold interpretation warrants caution.

The robustness of this outcome was examined in two further analyses. Adjusting the post-brushing level for the pre-brushing level gave the same group effect ($F(3,35) = 13.25$; $p < 0.001$), but Ra at T1 carried almost no predictive information for Ra at T2 ($F = 0.24$; $p = 0.627$), so the adjusted analysis is close to the unadjusted comparison of T2 levels; the homogeneity-of-slopes test was itself marginally significant ($p = 0.0499$), so that assumption is not securely met, and the adjusted analysis is reported as supportive rather than definitive. In that analysis, RUW + RPP separated from the control (+0.089 µm; 95% CI +0.043 to +0.135 µm), which it did not on the primary phase-specific outcome—a difference that follows from the higher roughness with which that group entered the brushing phase rather than from what happened during it. Both results are reported, and statements elsewhere that the prophylaxis paste was the only application to roughen the surface refer to the primary ΔRa outcome. Covariate overlap was adequate for this adjustment: the four groups share a common Ra(T1) range of 0.085–0.130 µm, no pairwise comparison is disjoint and the adjustment point lies inside every group's observed range, so no extrapolation is involved. The borderline contrast between RUW + DPP and RUW depends on the variance assumption (Games–Howell $p = 0.058$; Tukey HSD $p = 0.020$) and is reported as inconclusive.

3.2. Staining and Stain Removal

The colour change produced by the coffee solution (ΔE01) differed among groups ($F(3,36) = 12.25$; $p < 0.001$; $\eta^2 = 0.505$; Table 3, Figure 3). RUW + RPP stained significantly more than the other three groups ($p \leq 0.003$), which did not differ from one another. Baseline lightness did not differ ($p = 0.119$), so this was not caused by the initial colour. At group level, however, the group with the highest pre-staining Ra also stained most and the group with the lowest stained least, in line with the established relationship between roughness and stain uptake. The shift was not driven by a few specimens: all ten

RUW + RPP specimens exceeded the 75th percentile of the other groups. Stain removal was therefore analysed by covariance analysis.

Table 3. CIEDE2000 colour differences (ΔE00).

Group	ΔE01 (T0 → T1)	ΔE12 (T1 → T2)	ΔE12 Adjusted *	ΔE02 (T0 → T2)	ΔE02 < 1.8
CG	3.63 ± 0.94 b	1.91 ± 0.68 c	2.26	2.07 ± 0.42 a	2/10
RUW	4.00 ± 0.78 b	2.78 ± 0.92 b	2.84	1.52 ± 0.32 b	9/10
RUW + RPP	5.23 ± 0.69 a	3.71 ± 0.69 a	2.79	2.20 ± 0.23 a	0/10
RUW + DPP	3.43 ± 0.42 b	2.21 ± 0.47 bc	2.72	2.06 ± 0.33 a	3/10
p	<0.001	<0.001	0.020	<0.001	—

Mean ± standard deviation. Groups with different letters within a column differ significantly (Tukey HSD, $p < 0.05$). * Means adjusted by analysis of covariance with ΔE01 as covariate.

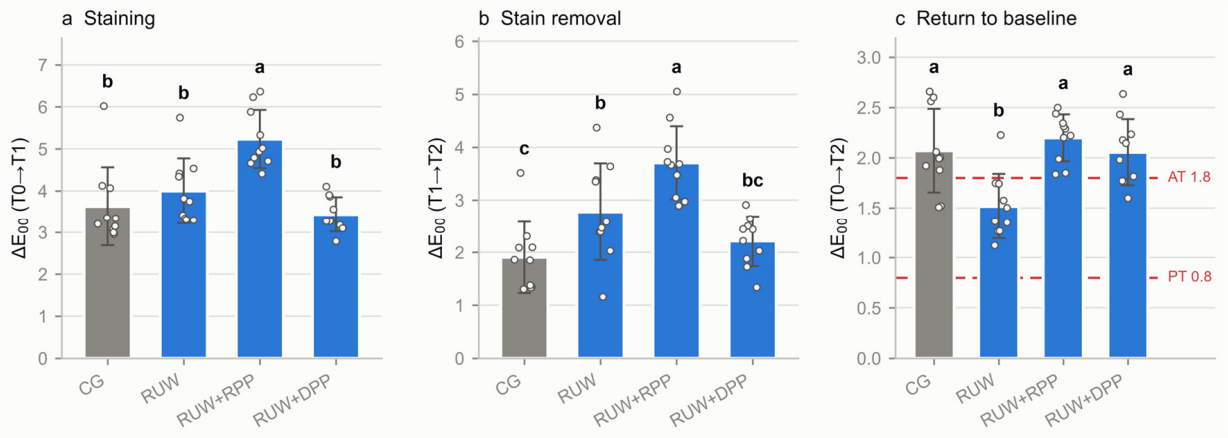


Figure 3. CIEDE2000 colour differences. (**a**) Staining; (**b**) colour removed by brushing; (**c**) difference remaining from baseline. Dashed lines in (**c**) mark the AT = 1.8 and PT = 0.8 thresholds; different letters denote significance (Tukey HSD, $p < 0.05$).

Stain removal (ΔE12) also differed ($F(3,36) = 12.48$; $p < 0.001$; $\eta^2 = 0.510$), highest in RUW + RPP. With staining as covariate, however, the covariate effect was strong ($F = 70.90$; $p < 0.001$), the group effect weakened ($F = 3.75$; $p = 0.020$) and adjusted means converged between 2.26 and 2.84, slopes being homogeneous ($p = 0.070$). The apparent superiority of RUW + RPP therefore largely reflects its greater initial staining.

3.3. Return to Baseline Colour

The primary colour outcome, ΔE02, differed among groups ($F(3,36) = 8.32$; $p < 0.001$; $\eta^2 = 0.410$). RUW was significantly lower than the other three groups ($p \leq 0.005$), which did not differ from one another. Only the RUW mean fell below the acceptability threshold, with 9/10 specimens below it against 2/10 in CG, 3/10 in RUW + DPP and 0/10 in RUW + RPP. No specimen in any group fell below the perceptibility threshold: a visually detectable difference from baseline remained everywhere.

Read across the three primary outcomes, the toothpaste used alone was the only regimen whose mean residual difference from baseline fell below the acceptability threshold (ΔE00 1.52 ± 0.32, against 2.07 ± 0.42 for the untreated control), while leaving roughness indistinguishable from that control (ΔRa +0.038 ± 0.030 against +0.003 ± 0.040 µm; $p = 0.404$) and net whiteness at baseline (ΔWID −0.20 ± 0.81). Neither combined regimen improved on it: adding the home-use paste left a larger residual difference (2.20 ± 0.23), and adding the prophylaxis paste left both a larger residual difference (2.06 ± 0.33) and the

only measurable roughening. These are the values measured for the products tested here and cannot be extended to whitening toothpastes as a class.

3.4. Whiteness Index

The primary whitening outcome, net change from baseline (ΔWID02), differed among groups ($F(3,36) = 14.21$; $p < 0.001$; Table 4, Figure 4). RUW + DPP specimens became whiter than at baseline (+1.73 ± 0.72), whereas the others remained at or below theirs. Tukey HSD separated RUW + DPP from all three ($p \leq 0.002$), with no difference among the remainder ($p \geq 0.095$). Because baseline whiteness differed ($p < 0.001$), the comparison was repeated with WID(T0) as covariate; slopes were homogeneous ($p = 0.332$) and the group effect held ($F(3,35) = 7.16$; $p < 0.001$). Baseline whiteness and net change were negatively correlated ($r = -0.64$; $p < 0.001$), but this is an algebraic consequence of subtracting the baseline from the follow-up value and is not independent evidence of regression to the mean; the uncoupled test showed no detectable association between baseline and post-brushing whiteness ($r = -0.13$; $p = 0.415$).

Table 4. Whiteness index (WID) and change relative to two reference points.

Group	WID (T0)	WID (T1)	WID (T2)	ΔWID02 (T2 − T0) *	ΔWID12 (T2 − T1) †	ΔWID02 Adjusted ‡
CG	14.88 ± 0.73	10.85 ± 1.41	13.97 ± 0.94	−0.91 ± 1.25 b	3.12 ± 0.95 b	−0.63 b
RUW	14.94 ± 0.73	10.36 ± 2.36	14.73 ± 0.74	−0.20 ± 0.81 b	4.37 ± 2.65 b	+0.12 b
RUW + RPP	14.23 ± 0.39	6.61 ± 1.95	14.33 ± 1.01	+0.10 ± 0.87 b	7.72 ± 1.91 a	−0.08 b
RUW + DPP	13.90 ± 0.37	11.29 ± 0.76	15.63 ± 0.67	+1.73 ± 0.72 a	4.34 ± 1.23 b	+1.31 a
p	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Mean ± standard deviation; different letters within a column denote significance (Tukey HSD, $p < 0.05$). * Primary whitening outcome. † Secondary outcome; it includes the T1 reading, is coupled to the degree of staining and is not a direct measure of whitening efficacy. ‡ Estimated from analysis of covariance with WID(T0) as covariate and evaluated at the overall mean baseline whiteness (14.49); the group effect remains significant ($F(3,35) = 7.16$; $p < 0.001$), and slopes are homogeneous ($p = 0.332$). All pairwise contrasts with Tukey-corrected 95% confidence intervals are given in the Supplementary Materials.

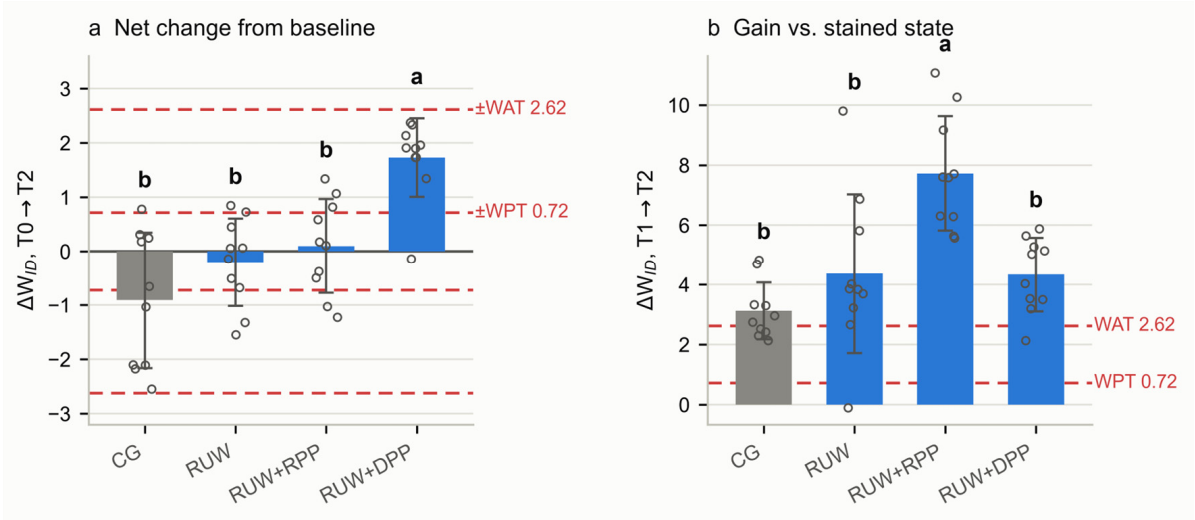


Figure 4. Whiteness index. (**a**) Primary outcome: net change from baseline (ΔWID02); only RUW + DPP rose above baseline. (**b**) Secondary outcome: gain relative to the stained state (ΔWID12), coupled to the degree of staining. Dashed lines are the whiteness perceptibility (WPT = 0.72) and acceptability (WAT = 2.62) thresholds; in (**a**), where the quantity is a signed change, they are drawn symmetrically at ±0.72 and ±2.62. Bars are means, whiskers standard deviations, circles individual specimens; letters denote significance (Tukey HSD, $p < 0.05$) within each panel.

As a secondary outcome, gain relative to the stained state (ΔWID12) also differed ($F(3,36) = 11.93$; $p < 0.001$; $\eta^2 = 0.499$), highest in RUW + RPP. This measure includes the T1 reading and is therefore coupled to the degree of staining; once post-brushing whiteness was adjusted for baseline whiteness and group, no detectable relationship with staining remained ($p = 0.102$). Across the three time points, the main effect of time ($F(2,72) = 239.48$; $p < 0.001$) and the group × time interaction ($F(6,72) = 12.49$; $p < 0.001$) were significant. In paired comparisons, the control ended below baseline (−0.91 [−1.80, −0.01]), whereas for RUW (−0.20 [−0.78, +0.38]) and RUW + RPP (+0.10 [−0.53, +0.72]), no departure from baseline was detected, with intervals reaching the whiteness perceptibility reference value, so a perceptible departure cannot be excluded for either arm; RUW + DPP ended above baseline (+1.73 [+1.21, +2.25]).

3.5. CIELAB Coordinates and the Source of the Whiteness Change

The underlying coordinates are given in Table 5. Mean L* fell below baseline in every group at T2 (ΔL* −2.84 to −1.99), so no group became lighter than it had been at baseline, and the differences in WID arise mainly from the chromatic terms. Decomposing ΔWID02 into its three components shows that the whole of the +1.73 gain in RUW + DPP is carried by the b* term (+3.16), while the L* term is negative (−1.15) and the a* term small (−0.28); the corresponding b* contributions were +1.35 in CG, +1.40 in RUW and +2.26 in RUW + RPP. Yellowness fell furthest in RUW + DPP (b* 27.86 → 24.99). Because ΔE00 is unsigned, it carries no direction: the colour removed during brushing (ΔE12) tracked the amount of staining ($r = +0.887$; $p < 0.001$) rather than the approach to baseline, and no specimen in any group moved past its baseline colour (0/40). Roughening and whiteness gain were associated across groups ($r = +0.437$; $p = 0.005$) but not within them (pooled within-group $r = -0.032$; $p = 0.843$). The individual specimen trajectories behind this pattern are shown in Figure 5.

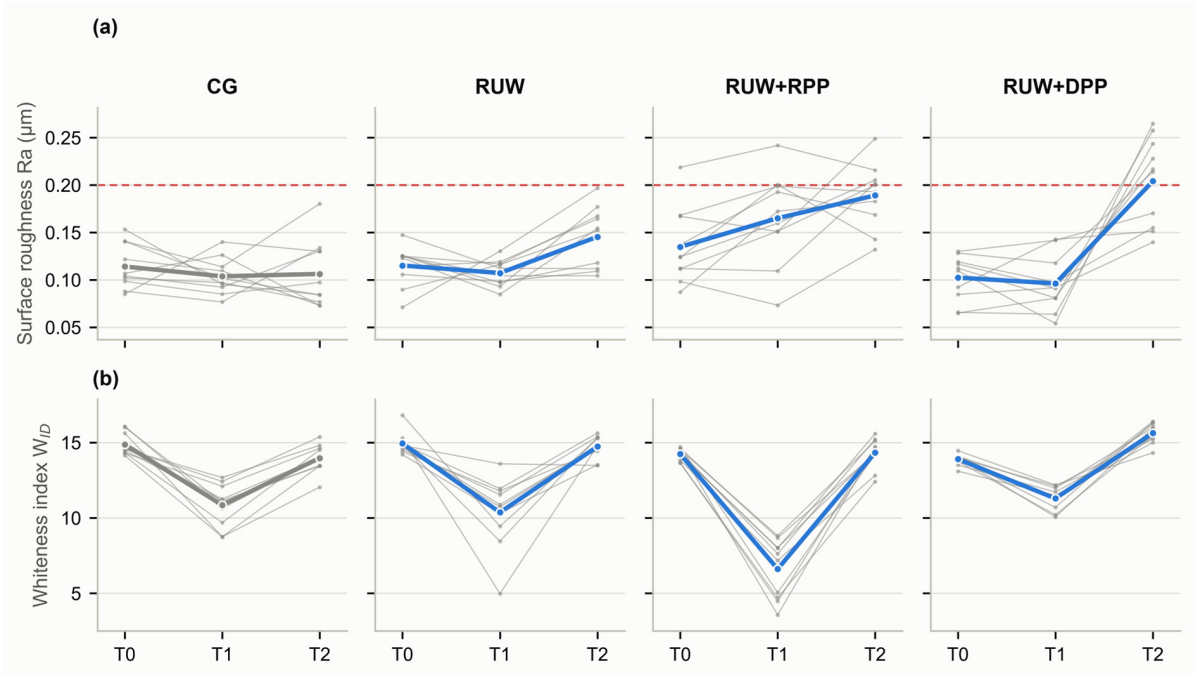


Figure 5. Individual specimen trajectories across the three time points. (**a**) Surface roughness; the dashed line is the 0.2 µm plaque-retention threshold. (**b**) Whiteness index. Thin grey lines are individual specimens ($n = 10$ per group), the heavy line the group mean. The divergence present before the group-specific intervention—visible in the RUW + RPP panel of (**a**) between T0 and T1 and in the depth of the T1 minimum in (**b**)—is unrelated to the treatment, which was applied only between T1 and T2.

Table 5. CIELAB coordinates at the three time points and change from baseline.

Group	Time	L*	a*	b*	ΔL*	Δa*	Δb*
CG	T0	81.44 ± 0.51	−1.29 ± 0.40	27.03 ± 0.68	—	—	—
	T1	76.38 ± 1.34	−0.97 ± 0.20	27.66 ± 0.72	−5.07	+0.32	+0.63
	T2	78.69 ± 0.48	−0.92 ± 0.14	25.80 ± 0.54	−2.75	+0.37	−1.23
RUW	T0	81.80 ± 0.61	−1.28 ± 0.18	27.13 ± 0.42	—	—	—
	T1	76.21 ± 1.07	−0.97 ± 0.24	28.04 ± 1.34	−5.60	+0.31	+0.91
	T2	79.81 ± 0.42	−1.03 ± 0.19	25.85 ± 0.44	−1.99	+0.25	−1.27
RUW + RPP	T0	82.03 ± 0.48	−1.22 ± 0.09	27.75 ± 0.30	—	—	—
	T1	74.78 ± 0.75	−0.58 ± 0.26	29.96 ± 0.95	−7.25	+0.64	+2.20
	T2	79.19 ± 0.58	−0.92 ± 0.14	25.70 ± 0.51	−2.84	+0.31	−2.06
RUW + DPP	T0	81.73 ± 0.52	−1.20 ± 0.11	27.86 ± 0.29	—	—	—
	T1	76.84 ± 0.33	−0.99 ± 0.10	27.52 ± 0.55	−4.90	+0.21	−0.35
	T2	79.48 ± 0.35	−1.08 ± 0.11	24.99 ± 0.41	−2.25	+0.12	−2.87

Mean ± standard deviation, $n = 10$ per group. Δ values are the change from baseline (T0). Coordinates were recovered from the original instrument readings and reproduce the published ΔE00 values exactly for all 40 specimens and the published WID values to within 0.0002 units; the specimen-level file is provided as Supplementary Material (Table S1).

4. Discussion

All parameters differed significantly among groups, and the null hypothesis was rejected. The most striking finding is that, on the primary roughness outcome, the professional prophylaxis paste was both the only application to roughen the surface measurably during the brushing phase and the only one to leave it whiter than baseline, whereas the added home-use polishing paste was indistinguishable from the toothpaste alone in either respect. On residual colour difference from baseline, however, the added home-use paste performed significantly worse than the toothpaste alone.

During abrasive procedures, the organic matrix is generally removed first and filler particles protrude, producing a rougher profile. This explains the roughening seen even in the control group: brushing with water alone has been reported to produce a measurable increase [9], and Gömleksiz and Okumuş [15] likewise found that the control group roughened while retaining the lowest value, in agreement with our data.

The direction of the effect is not the same in every design. Oliveira et al. [5] applied a simulation corresponding to two years of use and found roughness reduced in all groups, attributing this to the polishing effect of prolonged brushing, with a prophylaxis paste producing no additional roughening. Time scale may therefore be decisive, a one-month exposure bringing the abrasive effect to the fore. Closer to our design, Da Rosa et al. [36] tested three whitening toothpastes of clearly different abrasivity on a nanohybrid composite over one simulated month and found no significant increase. Our findings agree only in part: the toothpaste alone was likewise indistinguishable from the control, but the prophylaxis paste roughened measurably—a difference pointing to product class, since all products compared there were home-use toothpastes.

Meseli et al. [17] tested the same home-use polishing paste and found no adverse effect on roughness. Three differences may explain this. The paste was applied once a week there and twice a week here, the upper end of the manufacturer's recommendation. Their groups held seven specimens, too few to detect a difference of the size seen here; that result therefore means the difference could not be detected, not that none exists. Their brushing

simulator also held force and stroke more tightly than our set-up, and they diluted the paste with artificial saliva, which may change how the abrasive particles meet the surface.

For the prophylaxis paste (RDA 150), the literature is internally contradictory. Liebermann et al. [4] found it among the groups producing the lowest roughness, but it was applied in two steps there, and the reported value belongs to a surface finished with a less abrasive step, on an indirect CAD/CAM composite. Arkan et al. [7] tested the same product on a resin nanoceramic and found the greatest increase in that group; our findings agree with the latter. Prophylaxis pastes can roughen measurably even in a single application: Lacerda Neto et al. [6], using the same profilometer, reported that 30 s of pumice paste markedly raised roughness on a nanohybrid composite, with a significant material × prophylaxis interaction that led the authors to stress material dependence [8].

Clinically, the mean roughness of the prophylaxis paste group reached the 0.2 µm plaque-retention threshold [3], above which plaque accumulation increases, though its confidence interval includes that value and the other arms stayed below it. Lacerda Neto et al. [6] recommend re-polishing surfaces after professional prophylaxis. The topographic change observed here is consistent with the situation that recommendation addresses, but no re-polishing step was applied in this study, so the present data can neither confirm nor quantify its benefit.

The largest roughening occurred with the product carrying the higher of the two literature RDA values, but the four protocols cannot be ranked by abrasivity: no RDA is available for the whitening toothpaste common to all three treatment arms (Table 1). RDA is in any case a standardised measure of abrasion against dentine rather than a measurement of wear on a composite, and is used here only as contextual product information. Meseli et al. [17] found no difference in composite surface roughness between toothpastes of graded RDA, and Kurucu Karadeniz et al. [9] reported that a toothpaste of lower RDA roughened more than one of higher RDA, emphasising particle structure, shape and size; Liebermann et al. [4] showed by electron microscopy that round, small particles leave smoother surfaces. Here, the paste reported at RDA 150 roughened significantly more than the paste reported at approximately RDA 105 despite having about one fifth of the application time, which abrasivity alone does not explain. Since roughening cannot be assumed to progress linearly, this comparison was assessed in direction only.

On the colour side, return to baseline was chosen as the primary outcome rather than the amount of stain removed, and the findings justify that choice: RUW + RPP stained significantly more and stood out in the raw stain-removal values, but once staining was taken as a covariate, the group effect weakened and adjusted means converged. The hierarchy in the raw data largely reflects the initial difference in staining.

Only the toothpaste alone remained below the acceptability threshold for return to baseline colour, consistent with the mixed picture in the literature. Meseli et al. [17] reported colour difference above the threshold in all groups after one month, whereas Kurucu Karadeniz et al. [9] found some groups below it, and Yıldırım and Recen [37] reported that a fine-grained prophylactic paste brought it below their threshold. Variation in material, staining protocol and threshold selection may explain these differences [22], and the use of different thresholds limits direct numerical comparison.

Two studies adopting our thresholds show that the outcome depends largely on the material: Kaplan et al. [16] found the main effect of composite type significant but that of toothpaste type not, and Celik and İşcan Yapar [1] reported a nanohybrid material to be least affected by both coffee and brushing. That a single nanohybrid composite was tested here should be read in this light, since differences may be larger in less stable systems. The material was introduced to the market only recently, and no independent peer-reviewed data on its behaviour under staining and brushing could be identified. That

scarcity limits the context in which the present findings can be placed and is a further reason for caution in generalising them; it is not offered as a justification for the study. The material was supplied by its manufacturer, but the study was designed, conducted, analysed and reported independently, and the manufacturer had no part in it.

No specimen fell below the perceptibility threshold. This fits the view that coffee causes subsurface as well as surface discolouration: Mehrgan et al. [2] note that its polar, slowly released pigments are absorbed by the composite, and Mocny-Pachońska et al. [21] reported clinically that no paste removed stains completely. Since mechanical methods cannot reach subsurface discolouration, this is expected.

The whiteness index carries different information from colour difference: ΔE00 measures perceptual distance without direction, whereas WID expresses whiteness itself, which makes the reference point critical. In gain relative to the stained state, all groups exceeded the whiteness acceptability threshold, and the highest value appeared in the group that stained most—unavoidable, since this measure cannot separate the stain to be removed from whitening capacity. Net change from baseline gives a different picture, only the prophylaxis paste group ending whiter than baseline. Decomposing that change shows where it came from: the whole of the +1.73 gain in RUW + DPP is carried by the b* term (contribution +3.16), while the L* term is negative (−1.15). Mean L* fell below baseline in every group, so no arm became lighter; the prophylaxis paste group became less yellow (b* 27.86 → 24.99). A whiteness gain on the WID scale is therefore not the same as a lighter specimen, and the two should not be described interchangeably.

Studies using WID are few. Yazkan et al. [18] compared six toothpastes on a stained anterior composite using our thresholds and the same spectrophotometer family, the highest change occurring with a high-abrasivity toothpaste, although absolute values are not comparable because that material started less white. Arkan et al. [7] reported that the difference in the Detartrine group exceeded clinical acceptability. Most studies remain limited to ΔE00, which cannot alone state whether a specimen became whiter; reporting WID alongside it fills that gap for products making a whitening claim.

Read together, the findings give the clearest implication of this study. On the primary roughness outcome, the prophylaxis paste was the only protocol to roughen the surface measurably, and it was also the only one to leave a surface whiter than baseline, its post-brushing mean reaching the plaque-retention threshold—both from only 12 s of application. The two findings occur in the same arm, but the present endpoints do not establish a causal link between them. Across specimens, the association between roughening and whiteness gain is entirely a between-group effect: pooled within groups, the correlation disappears ($r = -0.03$; $p = 0.843$), so specimens that roughened more did not become whiter. Resin loss, filler exposure and the depth of pigment penetration were not measured, and the surface-layer explanation therefore remains untested. The home-use paste, by contrast, was applied more than five times as long yet could not be distinguished from the toothpaste alone in roughness change or net whiteness change; on residual colour difference from baseline, it performed significantly worse (ΔE02 2.20 ± 0.23 versus 1.52 ± 0.32), although that arm had also stained more heavily before treatment (ΔE01 5.23 ± 0.69 versus 4.00 ± 0.78), so the difference cannot be attributed to the paste itself. Its apparent superiority in gain relative to the stained state arose from that same greater staining.

Three practical implications follow, each within the limits set out below. Adding a home-use polishing paste to a whitening toothpaste produced no detectable benefit in any outcome of this study, and on residual colour difference from baseline, the combined regimen performed significantly worse than the toothpaste alone; the design does not allow that difference to be attributed to the paste itself, but nothing in these data supports adding one. Because the prophylaxis paste produced a clear whiteness gain while roughening

the surface, re-polishing after professional prophylaxis is a strategy that deserves direct evaluation; it was not tested here. Moreover, since no protocol reduced staining below the perceptibility threshold and mechanical methods cannot reach subsurface discolouration, colour recovery of a heavily stained restoration may remain incomplete—whether replacement is preferable was not tested in this study.

Several limitations apply. The most important concerns the roughness measurements: the three readings per specimen were taken from three different areas, so their dispersion characterises the spatial heterogeneity of the surface together with measurement error rather than instrument repeatability alone. That dispersion is comparable to the differences between group means, which is why baseline roughness does not predict a specimen's post-brushing value and change-score analysis loses power. The groups also diverge before the intervention in baseline whiteness, degree of staining and roughness at T1, the last arising entirely from the RUW + RPP arm, although all specimens underwent identical procedures to that point. The allocation was intended to be random, but the method was not recorded, and that gap has not been filled with an assumed procedure: whether the imbalance reflects chance at allocation or variability in specimen production cannot be determined. What the record does establish is reported in Section 2.2. The roughness findings should therefore be read as comparative and exploratory rather than causal, and a confirmatory study should use documented allocation balanced on pretreatment measurements, with processing and measurement order independently randomised and recorded and with a blinded assessor.

Two further points concern the colour analyses. Baseline whiteness and net whiteness change are negatively correlated, but that correlation follows algebraically from subtracting the baseline and is not by itself evidence of regression to the mean; the whitening analyses were nevertheless repeated with baseline whiteness as covariate and with the immediately preceding reading as covariate, and the group effect held in both. Covariate overlap was inspected: for roughness, the four groups share a common range and the adjustment involves no extrapolation; for baseline whiteness, the common range is narrow and the adjustment point lies just outside it (by 0.03 WID units), which is negligible; but for WID(T1) and for $\Delta E01$, the four-group intersection is empty, and the adjusted estimates for the arms at the extremes rest partly on extrapolation, so those models are supportive rather than decisive. This is one reason why the primary colour outcome was chosen to need no covariate. Finally, the study was powered for large effects only; although all significant differences exceeded that threshold, type II error cannot be excluded below it, and non-significant comparisons are reported as differences not detected rather than as equivalence, for which no margin was specified.

This study was conducted in vitro. Salivary flow and buffering, the acquired pellicle, microbial colonisation, pH and temperature fluctuations and brushing habits could not be reproduced, and the flat specimens do not reflect clinical geometry. The protocol represents an assumed one month of cumulative use [17], and no thermocycling was applied. Paste application times were unequal by design, reflecting how the products are actually used: regular home use versus a single short clinical application. The comparison is therefore between complete regimens; because the added applications differ in duration and in the way the product is delivered, the design does not isolate the effect of a paste formulation from that of the additional mechanical exposure, and a time-matched additional-brushing control would be needed to do so. The arms also differ in class, R.O.C.S. Pro Polishing being a home-use paste classified among toothpastes by Meseli et al. [17]; the two paste arms therefore cannot be compared directly, although comparison of each against the control and the toothpaste alone is unaffected. Surface assessment was limited to Ra, without gloss, microscopy or any measure of material loss. This is not a minor omission: surfaces leaving almost identical Ra can differ markedly in gloss [4], and the two properties

vary independently with force and time [20], so whether the rougher surface left by the prophylaxis paste was better or worse in overall appearance cannot be decided from these data. Gloss, together with a direct measure of material loss, is accordingly a primary addition for any confirmatory study. The specimens exist only in their final, brushed condition, and no surface record was preserved at baseline or after staining, so paired before-and-after microscopy cannot now be produced for this set; testing the surface-layer explanation requires a new experiment designed for it. Only one staining medium was used, so the behaviour of other dietary colourants and the sedimentation of the insoluble fraction present in a brew from ground beans were not modelled. No dental hard tissue was tested: every specimen was a composite disc, so nothing in these data bears on how either paste affects enamel or dentine, and the abrasivity values quoted in Table 1 are dentine-based measures that cannot be transferred to the composite response. Finally, a single composite and shade were tested, so the results should not be generalised [6,16], and the spectrophotometer is a clinical device intended for natural teeth.

5. Conclusions

Within the limits of this in vitro study, the professional prophylaxis paste was the only application that roughened the surface measurably during the brushing phase and the only one that left it whiter than baseline; its mean post-brushing roughness reached the 0.2 μm plaque-retention threshold, although the confidence interval includes that value and three of four arms stayed below it; re-polishing afterwards is a strategy that requires direct evaluation. Adding a home-use polishing paste could not be distinguished from the toothpaste alone in roughness change or net whiteness change, but left a significantly larger residual colour difference from baseline; because that arm was also the more heavily stained before treatment, the difference cannot be attributed to the paste itself. Only the toothpaste alone remained below the acceptability threshold for colour recovery, and no protocol reduced staining below the perceptibility threshold. Whiteness gain and residual colour mismatch favoured different arms and should be read as complementary outcomes rather than as a single ranking.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/app162010139/s1, Table S1: Specimen-level dataset (file Table_S1_specimen_level.csv); Table S2: Individual instrument readings (file Table_S2_individual_readings.csv); File S1: Description of the data files and of the provenance of the colour coordinates (file README.txt); File S2: Syntax of the original analysis (file SPSS_syntax.sps); File S3: Python scripts for the covariate-adjusted analyses and the figures (folder analysis_scripts); File S4: Outputs of those scripts (folder statistical_output).

Author Contributions: Conceptualisation, M.Ş. and Z.H.K.; methodology, M.Ş. and E.A.; formal analysis, M.Ş. and A.A.A.; investigation, E.A.; resources, E.E.K.; data curation, E.A.; writing—original draft preparation, M.Ş.; writing—review and editing, M.Ş., E.A., A.A.A., E.E.K. and Z.H.K.; visualisation, M.Ş.; supervision, M.Ş., A.A.A., E.E.K. and Z.H.K.; project administration, M.Ş. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding. The composite resin tested (Elóra APS) was supplied free of charge by its manufacturer, FGM Dental Group (Joinville, SC, Brazil); all other materials were purchased by the authors at their own expense.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data supporting the conclusions of this article are provided as Supplementary Materials. Table S1 gives the specimen-level dataset (surface roughness, CIELAB

coordinates and every derived endpoint for each of the 40 specimens at each time point) and Table S2 the individual instrument readings from which those values were computed, with specimen identifiers. No specimen was excluded, and no reading was discarded. The syntax of the original analysis (File S2), the Python scripts used for the covariate-adjusted analyses and the figures (File S3) and the corresponding outputs (File S4) are included, together with File S1, which describes the files and the provenance of the colour coordinates.

Acknowledgments: During the preparation of this manuscript, the authors used Claude (Anthropic, San Francisco, CA, USA; Opus 5 and Opus 5.5) for the purposes of cross-checking the statistical analyses, recovering the CIELAB coordinates from the original instrument records, translating and language-editing the text, producing the figures as scripted plots of the study's own measured data and preparing the Supplementary Files. No AI tool was used to generate, alter or fabricate primary data, specimen images or results. No generative image model was used, no figure content was synthesised and no photographic material is included. All content produced with the aid of this technology was fully reviewed, edited and validated by the authors. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest. The composite resin tested was supplied free of charge by its manufacturer, FGM Dental Group. The donor had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

ANOVA	analysis of variance
AT	acceptability threshold for colour difference (ΔE00 = 1.8)
CG	control group (distilled water)
CIEDE2000	CIE 2000 colour-difference formula
CIELAB	CIE 1976 L*a*b* colour space
DPP	professional prophylaxis paste
PT	perceptibility threshold for colour difference (ΔE00 = 0.8)
Ra	arithmetic mean surface roughness
RDA	relative dentin abrasivity
RPP	home-use polishing paste
RUW	whitening toothpaste
T0, T1, T2	baseline, after staining, after brushing
WAT	whiteness acceptability threshold (2.62 WID units)
WID	whiteness index for dentistry
WPT	whiteness perceptibility threshold (0.72 WID units)
ΔE00	CIEDE2000 colour difference

Appendix A

Table A1. Analysis-of-variance tables for the three primary outcomes.

Outcome	Source	SS	df	*F*	*p*	η^2
ΔRa (T2 − T1)	Group	0.0624	3	8.12	<0.001	0.404
	Residual	0.0922	36	—	—	—
	Shapiro–Wilk on residuals	W = 0.9840			0.832	
ΔE00 (T0 → T2)	Group	2.7469	3	8.32	<0.001	0.410
	Residual	3.9596	36	—	—	—
	Shapiro–Wilk on residuals	W = 0.9760			0.544	

Table A1. *Cont.*

Outcome	Source	SS	df	*F*	*p*	η^2
ΔWID (T2 − T0), WID(T0) as covariate	Group	15.4903	3	7.16	<0.001	0.329
	WID_T0	6.3401	1	8.79	0.005	0.135
	Residual	25.2423	35	—	—	—
	Shapiro–Wilk on residuals	W = 0.9591			0.156	

Type II sums of squares. η^2 is the proportion of the total sum of squares. Plots of residuals against the order of specimen processing or measurement could not be produced: neither the fabrication and polishing order nor the measurement order was recorded during the experiment, and no order has been reconstructed retrospectively (Section 2.2).

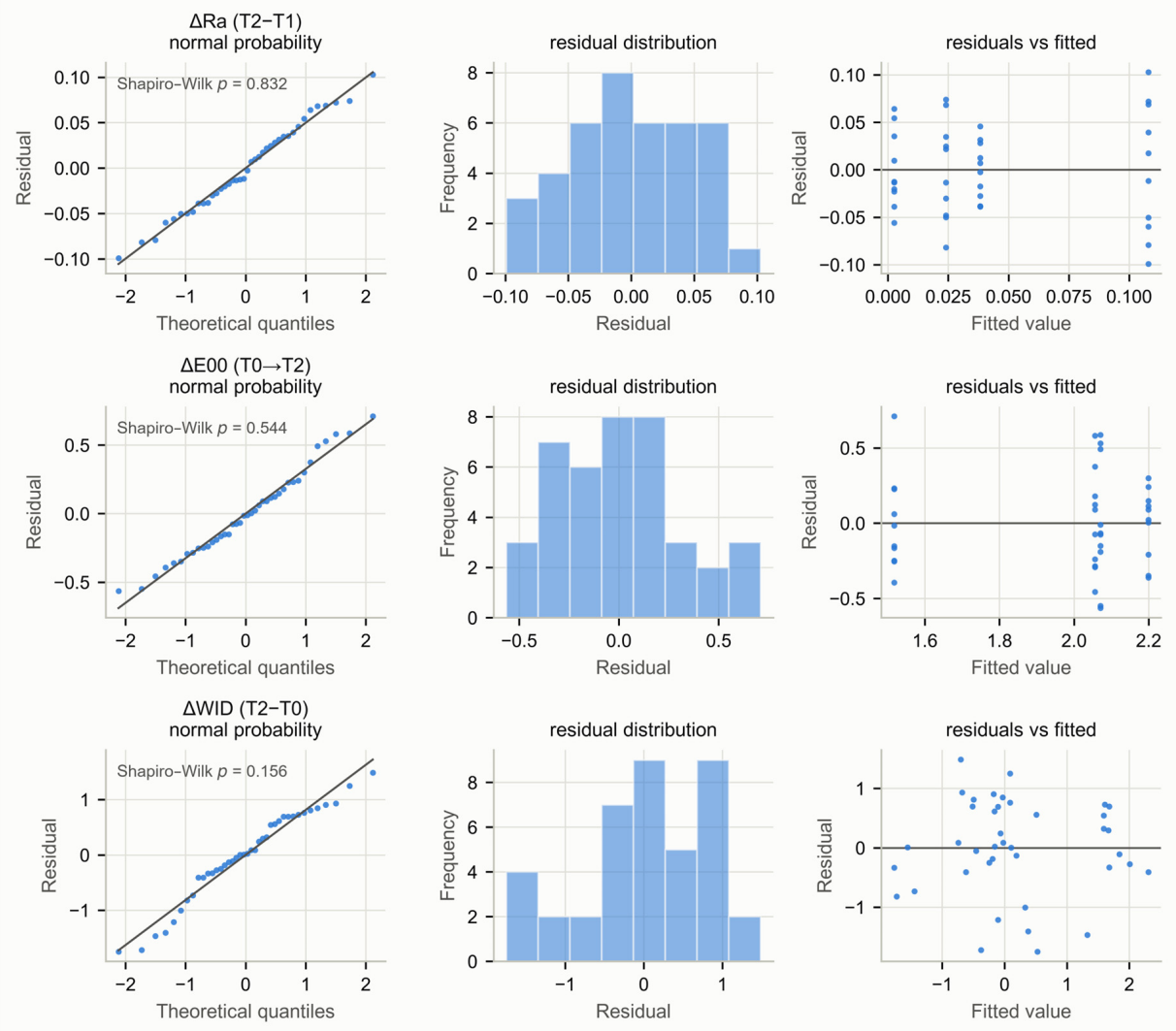


Figure A1. Model diagnostics for the three primary outcomes. (**Left**), normal probability plot of the residuals with the Shapiro–Wilk *p* value; (**centre**), residual distribution; (**right**), residuals against fitted values.

References

1. Celik, N.; Iscan Yapar, M. Colour stability of stained composite resins after brushing with whitening toothpaste. *Int. J. Dent. Hyg.* **2021**, *19*, 413–420. [CrossRef]
2. Mehrgan, S.; Kermanshah, H.; Omrani, L.R.; Ahmadi, E.; Rafeie, N. Comparison the effect of charcoal-containing, hydrogen peroxide-containing, and abrasive whitening toothpastes on color stability of a resin composite; an in vitro study. *BMC Oral Health* **2021**, *21*, 594. [CrossRef]

3. Bollen, C.M.L.; Lambrechts, P.; Quirynen, M. Comparison of surface roughness of oral hard materials to the threshold surface roughness for bacterial plaque retention: A review of the literature. *Dent. Mater.* **1997**, *13*, 258–269. [CrossRef]
4. Liebermann, A.; Spintzyk, S.; Reymus, M.; Schweizer, E.; Stawarczyk, B. Nine prophylactic polishing pastes: Impact on discoloration, gloss, and surface properties of a CAD/CAM resin composite. *Clin. Oral Investig.* **2019**, *23*, 327–335. [CrossRef]
5. Oliveira, R.S.; Simões, H.B.; Odilon, N.N.; Lima, M.J.P.; Campos, E.J. Effect of whitening dentifrices containing optical agent on the variation of color, roughness and mass of a nanoparticulate composite resin. *Rev. Odontol. UNESP* **2020**, *49*, e20200073. [CrossRef]
6. Lacerda Neto, J.; Castro González, J.A.; Costa, M.P.; Balardini, L.A.; Oliveira, M.F.; Amaral, M.; Zanatta, R.F. Surface roughness of resin-based composites subjected to professional prophylaxis. *Rev. Fac. Odontol. Porto Alegre* **2023**, *64*, e130840. [CrossRef]
7. Arkan, B.; Iscan, I.; Tinastepe, N. The impact of prophylactic polishing pastes on the surface roughness, color, and stain resistance of CAD/CAM restorative materials. *Dent. J.* **2025**, *13*, 212. [CrossRef]
8. Yurdaguven, H.; Aykor, A.; Ozel, E.; Sabuncu, H.; Soyman, M. Influence of a prophylaxis paste on surface roughness of different composites, porcelain, enamel and dentin surfaces. *Eur. J. Dent.* **2012**, *6*, 1–8. [CrossRef]
9. Kurucu Karadeniz, B.K.; Özdemir, S.B.; Özdemir, B.; Kaplan, F. Effect of different whitening toothpaste on surface roughness and discoloration of universal composite resins. *J. Turk. Dent. Res.* **2024**, *3*, 398–408. [CrossRef]
10. Sharma, G.; Wu, W.; Dalal, E.N. The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Res. Appl.* **2005**, *30*, 21–30. [CrossRef]
11. Ghinea, R.; Pérez, M.M.; Herrera, L.J.; Rivas, M.J.; Yebra, A.; Paravina, R.D. Color difference thresholds in dental ceramics. *J. Dent.* **2010**, *38*, e57–e64. [CrossRef]
12. Paravina, R.D.; Ghinea, R.; Herrera, L.J.; Della Bona, A.; Igiel, C.; Linninger, M.; Sakai, M.; Takahashi, H.; Tashkandi, E.; Pérez, M.M. Color difference thresholds in dentistry. *J. Esthet. Restor. Dent.* **2015**, *27*, S1–S9. [CrossRef]
13. Pérez, M.M.; Ghinea, R.; Rivas, M.J.; Yebra, A.; Ionescu, A.M.; Paravina, R.D.; Herrera, L.J. Development of a customized whiteness index for dentistry based on CIELAB color space. *Dent. Mater.* **2016**, *32*, 461–467. [CrossRef]
14. Pérez, M.M.; Herrera, L.J.; Carrillo, F.; Pecho, O.E.; Dudea, D.; Gasparik, C.; Ghinea, R.; Della Bona, A. Whiteness difference thresholds in dentistry. *Dent. Mater.* **2019**, *35*, 292–297. [CrossRef]
15. Gömleksiz, S.; Okumuş, Ö.F. The effect of whitening toothpastes on the color stability and surface roughness of stained resin composite. *BMC Oral Health* **2024**, *24*, 860. [CrossRef]
16. Kaplan, E.; Dündar, A.; Barutçugil, Ç. Color stability and surface roughness of esthetic resin composites following simulated toothbrushing with whitening toothpastes. *Odovtos Int. J. Dent. Sci.* **2026**, *28*, 154–167. [CrossRef]
17. Meseli, S.; Alkan, E.; Korkut, B.; Kanar, O.; Tagtekin, D. Abrasiveness and bleaching level of toothpastes on composite resins: A quantitative analysis using a novel brushing simulator. *Appl. Sci.* **2025**, *15*, 2314. [CrossRef]
18. Yazkan, B.; Yilmaz, E.; Yenidünya, Ö.G.; Akgül, N. Effects of different toothpastes on the color and whiteness of stained anterior composite resin. *J. Oral Sci.* **2023**, *65*, 246–250. [CrossRef]
19. Kim, H.-Y. Analysis of variance (ANOVA) comparing means of more than two groups. *Restor. Dent. Endod.* **2014**, *39*, 74–77. [CrossRef]
20. Heintze, S.D.; Forjanic, M.; Rousson, V. Surface roughness and gloss of dental materials as a function of force and polishing time in vitro. *Dent. Mater.* **2006**, *22*, 146–165. [CrossRef]
21. Mocny-Pachońska, K.; Jaworowska-Baranowska, P.; Mackiewicz, A.; Masłowiec, I.; Stochel, Z.; Trzcionka, A.; Tanasiewicz, M. Clinical comparison of the effectiveness of three prophylactic toothpastes in removing dental stains. *J. Stomatol.* **2019**, *72*, 58–62. [CrossRef]
22. Paolone, G.; Formiga, S.; De Palma, F.; Abbruzzese, L.; Chirico, L.; Scolavino, S.; Goracci, C.; Cantatore, G.; Vichi, A. Color stability of resin-based composites: Staining procedures with liquids—A narrative review. *J. Esthet. Restor. Dent.* **2022**, *34*, 865–887. [CrossRef]
23. Bizhang, M.; Schmidt, I.; Chun, Y.-H.P.; Arnold, W.H.; Zimmer, S. Toothbrush abrasivity in a long-term simulation on human dentin depends on brushing mode and bristle arrangement. *PLoS ONE* **2017**, *12*, e0172060. [CrossRef]
24. Heintze, S.D.; Forjanic, M.; Ohmiti, K.; Rousson, V. Surface deterioration of dental materials after simulated toothbrushing in relation to brushing time and load. *Dent. Mater.* **2010**, *26*, 306–319. [CrossRef]
25. *ISO 4288:1996*; Geometrical Product Specifications (GPS)—Surface Texture: Profile Method—Rules and Procedures for the Assessment of Surface Texture. International Organization for Standardization: Geneva, Switzerland, 1996; Withdrawn on 20 December 2021.
26. *ISO 21920-3:2021*; Geometrical Product Specifications (GPS)—Surface Texture: Profile—Part 3: Specification Operators. International Organization for Standardization: Geneva, Switzerland, 2021.
27. Aljamhan, A.; Habib, S.R.; AlSarhan, M.A.; AlZahrani, B.; AlOtaibi, H.; AlSunaidi, N. Effect of finishing and polishing on the surface roughness of bulk fill composites. *Open Dent. J.* **2021**, *15*, 25–32. [CrossRef]

28. Kim-Pusateri, S.; Brewer, J.D.; Davis, E.L.; Wee, A.G. Reliability and accuracy of four dental shade-matching devices. *J. Prosthet. Dent.* **2009**, *101*, 193–199. [CrossRef]
29. Shapiro, S.S.; Wilk, M.B. An analysis of variance test for normality (complete samples). *Biometrika* **1965**, *52*, 591–611. [CrossRef]
30. Ghasemi, A.; Zahediasl, S. Normality tests for statistical analysis: A guide for non-statisticians. *Int. J. Endocrinol. Metab.* **2012**, *10*, 486–489. [CrossRef]
31. Brown, M.B.; Forsythe, A.B. Robust tests for the equality of variances. *J. Am. Stat. Assoc.* **1974**, *69*, 364–367. [CrossRef]
32. Welch, B.L. On the comparison of several mean values: An alternative approach. *Biometrika* **1951**, *38*, 330–336. [CrossRef]
33. Kruskal, W.H.; Wallis, W.A. Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc.* **1952**, *47*, 583–621. [CrossRef]
34. Games, P.A.; Howell, J.F. Pairwise multiple comparison procedures with unequal N's and/or variances: A Monte Carlo study. *J. Educ. Stat.* **1976**, *1*, 113–125. [CrossRef]
35. Lee, S.; Lee, D.K. What is the proper way to apply the multiple comparison test? *Korean J. Anesthesiol.* **2018**, *71*, 353–360, Correction in *Korean J. Anesthesiol.* **2020**, *73*, 572. [CrossRef]
36. da Rosa, G.M.; da Silva, L.M.; de Menezes, M.; do Vale, H.F.; Regalado, D.F.; Pontes, D.G. Effect of whitening dentifrices on the surface roughness of a nanohybrid composite resin. *Eur. J. Dent.* **2016**, *10*, 170–175. [CrossRef]
37. Yıldırım, B.; Recen, D. Color stability and translucency of two CAD-CAM restorative materials subjected to mechanical polishing, staining, and prophylactic paste polishing procedures. *EÜ Dişhek. Fak. Derg.* **2021**, *42*, 1–8. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.