OPEN ACCESS

Journal of Science & Cycling
Breakthroughs in Cycling & Triathlon Sciences

Original Article

Inter-rater variability in 2D kinematic cycling analysis using Kinovea®: a cross-sectional study with 53 bike fitting professionals in Brazil

Andrei Pereira Pernambuco [1,2], Miguel Victor Silva Sousa [1], Izabely Arantes Santos [1], Laura Fonseca Vaz Do Amaral [1], Thaiane Paula Lima Henrique [1], Gustavo Cardoso Vieira [2], Janaine Cunha Polese [3], Rafael Honório Pereira Alves [4], and Marcelo Rocha de Morais [5]

Received: 15 May 2025 **Accepted:** 1 September 2025 **Published:** 10 September 2023

[1] Centro Universitário de Formiga – MG (UNIFOR – MG), Formiga – MG, Brazil
[2] Universidade de Itaúna (UIT), Itaúna – MG, Brazil.
[3] Universidade Federal de Minas Gerais (UFMG), Belo Horizonte - MG, Brazil. d Instituto Federal de Educação, Ciência
[4] Tecnologia de Minas Gerais – Campus Formiga (IFMG – Formiga), Formiga – MG, Brazil
[5] Universidade Católica de Brasília, Brasília - DF, Brazil.

Correspondence

Andrei Pereira Pernambuco

Centro Universitário de Formiga – MG and Universidade de Itaúna

pernambucoap@ymail.com

Abstract

Cycling kinematic analysis plays a central role in the bike fitting process, directly influencing decisions related to performance, comfort, and injury prevention. Consistency of measurements across evaluators is therefore essential for ensuring reliable outcomes in both clinical and performance contexts. This cross-sectional inter-observer agreement study evaluated inter-examiner variability in two-dimensional (2D) kinematic measurements obtained with Kinovea® software. A sample of 53 professional bike fitters from different regions of Brazil analyzed the same 40-second video of a cyclist pedaling on a stationary mountain bike. Each participant independently selected frames and measured seven predefined joint and positional angles. Statistical analyses included descriptive measures, Shapiro–Wilk normality testing, bootstrap confidence intervals, one-sided chi-square variance tests with Holm corrections, bias and empirical limits of agreement, Brown–Forsythe tests of dispersion, and Fleiss' κ for categorical KOPS classification. The results showed notable inter-examiner variability, particularly for knee extension (CV = 6.2%), trunk flexion (CV = 5.8%), and plantar flexion (CV = 4.8%), which exceeded predefined tolerance thresholds of 2–4% of the mean. By contrast, hip flexion, knee flexion, and armpit angle showed greater consistency. Subgroup analyses revealed no significant effect of professional experience or software used on measurement variability. These findings highlight that, even under identical testing conditions, methodological differences among raters can substantially influence kinematic measurements. The study underscores the need for standardized protocols and structured training in 2D motion analysis to improve reliability in bike fitting practice and ensure safer, more effective adjustments for cyclists.

Keywords

Bicycling Ergonomics; 2D Kinematic Analysis; Measurement Reproducibility; Bike Fitting Variability

1 Introduction

The increasing popularity of cycling as both a sport and a recreational activity has drawn greater attention to factors influencing cyclist performance, comfort, and injury prevention (Robidoux, 2022). In this context, many cyclists opt for bike fitting, a systematic procedure designed to evaluate and optimize the interaction between the cyclist and the bicycle (Kotler, 2016).

Cycling kinematic analysis has been extensively investigated as a tool to optimize performance, improve comfort, and prevent injuries in cyclists. Previous research has shown that the use of high-speed cameras, laser alignment systems, and motion analysis software allows precise quantification of joint angles, body posture, and pedaling mechanics, which is essential for identifying biomechanical inefficiencies and potential sources of overload (Bini et al., 2019; Scoz et al., 2022). While three-dimensional systems provide comprehensive spatial data, two-dimensional approaches offer simpler and more accessible assessments, although they may present limitations in accuracy (Norris & Olson, 2011; Thewlis et al., 2013). Studies have indicated that lower limb alignment, saddle height, and handlebar position significantly influence pedaling efficiency and comfort, with inadequate adjustments increasing the risk of overuse injuries (Quesada et al., 2019; Bini & Priego-Quesada, 2022). Despite these advances, there is still no consensus on the reliability of low-cost two-dimensional tools, such as Kinovea®, when applied in professional bike fitting contexts. Although Kinovea® has been validated for joint angle measurement in different sports, including resistance training and gait analysis (Puig-Diví et al., 2019; Sañudo et al., 2016), little is known about the consistency of these measurements among different evaluators in cycling-specific tasks. This gap is relevant because inconsistent measurements may lead to divergent adjustment recommendations, potentially compromising the cyclist's performance, comfort, and musculoskeletal health (Bini et al., 2020; Gonzales et al., 2020).

It is well established that an examiner's proficiency in data acquisition and interpretation is essential for effective decision-making (Scoz et al., 2021). Despite advances in the use of two-dimensional analysis tools for bicycle fitting, no studies have systematically quantified inter-examiner variability in the measurement of joint angles during bike fitting using Kinovea®, even under standardized evaluation conditions. This gap is relevant because differences in measurements between professionals can lead to distinct interpretations and adjustments, directly impacting posture, comfort, and injury prevention in cyclists, as evidenced by studies that have linked the lack of standardization to inconsistent results in biomechanical analyses (Scoz et al., 2021; Bini & Quesada, 2022). Considering this scenario, the present study hypothesized that, even when analyzing the same video recording, there would be significant variability between evaluators, especially for angles whose anatomical landmark identification requires greater precision and familiarity with the tool.

Therefore, the aim of this study was to assess inter-examiner variability in two-dimensional kinematic measurements during a cycling motion analysis performed under the same assessment conditions using the Kinovea® software.

2 Material and Methods

2.1 Study Design and Ethics

The study, designed as a cross-sectional inter-observer agreement study, was initiated following approval of the research protocol by the Human Research Ethics Committee of Centro Universitário de Formiga (UNIFOR-MG), duly accredited by the National Research Ethics Commission (CONEP). The approval was registered under opinion number 5.406.719. All procedures adhered to the ethical standards outlined in the Declaration of Helsinki and its amendments.

2.2 Participants

The sample was obtained through convenience and non-probabilistic sampling, comprising 53 professional bike fitters. This number was defined based on a power analysis using a one-sided chi-square variance test (α = 0.05, df = 52), with practical tolerance limits set at 2–10% of the mean for each angle. Under these parameters, the sample provides approximately 80% power to detect a 25% increase in variance and over 99% power to detect a 50% increase, aligning with the study's objective of evaluating inter-examiner variability.

Professionals working in bike fitting across different regions of Brazil (South, Southeast, Midwest, Northeast, and North) were invited to participate in the study. Recruitment involved active outreach via WhatsApp groups, Instagram profiles, other social networks such as Facebook and LinkedIn, websites, and referrals from previously contacted professionals.

Inclusion criteria included: bike fitters working in Brazil; having access to email or another digital communication channel; regular use of image editing and/or motion analysis software; and voluntary agreement to participate by signing the Informed Consent Form (ICF). Exclusion criteria included: failure to meet the academic qualification requirement (being a licensed physiotherapist or occupational therapist); failure to sign the informed consent form; failure to submit the completed research form by the deadline; incomplete form responses; or failure to respond to follow-up contact attempts. Additionally, participants were excluded if they did not follow the analysis instructions, submitted incomplete or corrupted files, or provided measurements inconsistent with the required evaluation procedure. Participants who met the inclusion criteria but later withdrew from the study were also excluded. At the end of the recruitment process, 53 bike fitters, all of whom were physiotherapists or physical education professionals with 3 months to 20 years of clinical experience (mean 6.10 ± 4.61 years) remained in the study. All eligible participants then proceeded to the measurement phase, as described below.

2.3 Procedures

After a brief explanation of the study's objectives, risks, and benefits, participants received a 40-second video via email showing a cyclist pedaling a mountain bike (MTB) on a fluid roller (Figure 1). The video was recorded using a Logitech C920 camera positioned from a distance of three meters and 150 cm above the ground, with HD resolution (1280×720 pixels) and a frame rate of 15 frames per second, with proper leveling and plumb alignment. Participants also received an electronic form (Google Forms) containing an identification section and fields related to angular and positional parameters evaluated during a bike fitting session.

Figure 1. Image extracted from the video sent to the research participants. It is possible to observe the positioning of the bike on the fluid roller and the front wheel leveler, as well as the markers applied to the cyclist's body.

Participants were instructed to measure seven variables using KINOVEA® software: (A) ankle plantar flexion angle, (B) knee extension angle, (C) knee over pedal spindle (KOPS), (D) trunk flexion angle, (E) hip flexion angle, (F) knee flexion angle, and (G) reach angle (armpit). To minimize discrepancies in measurement standards, the authors provided a sequence of reference photographs along with the electronic form, illustrating the expected measurement technique for each variable of interest (see Figure 2). No specific training was provided to participants for this study. The selected variables are part of the standard 2D pedaling analysis protocol. Although all participants received the same video, the selection of frames to be analyzed was made by each bike fitter individually, replicating typical professional practice.

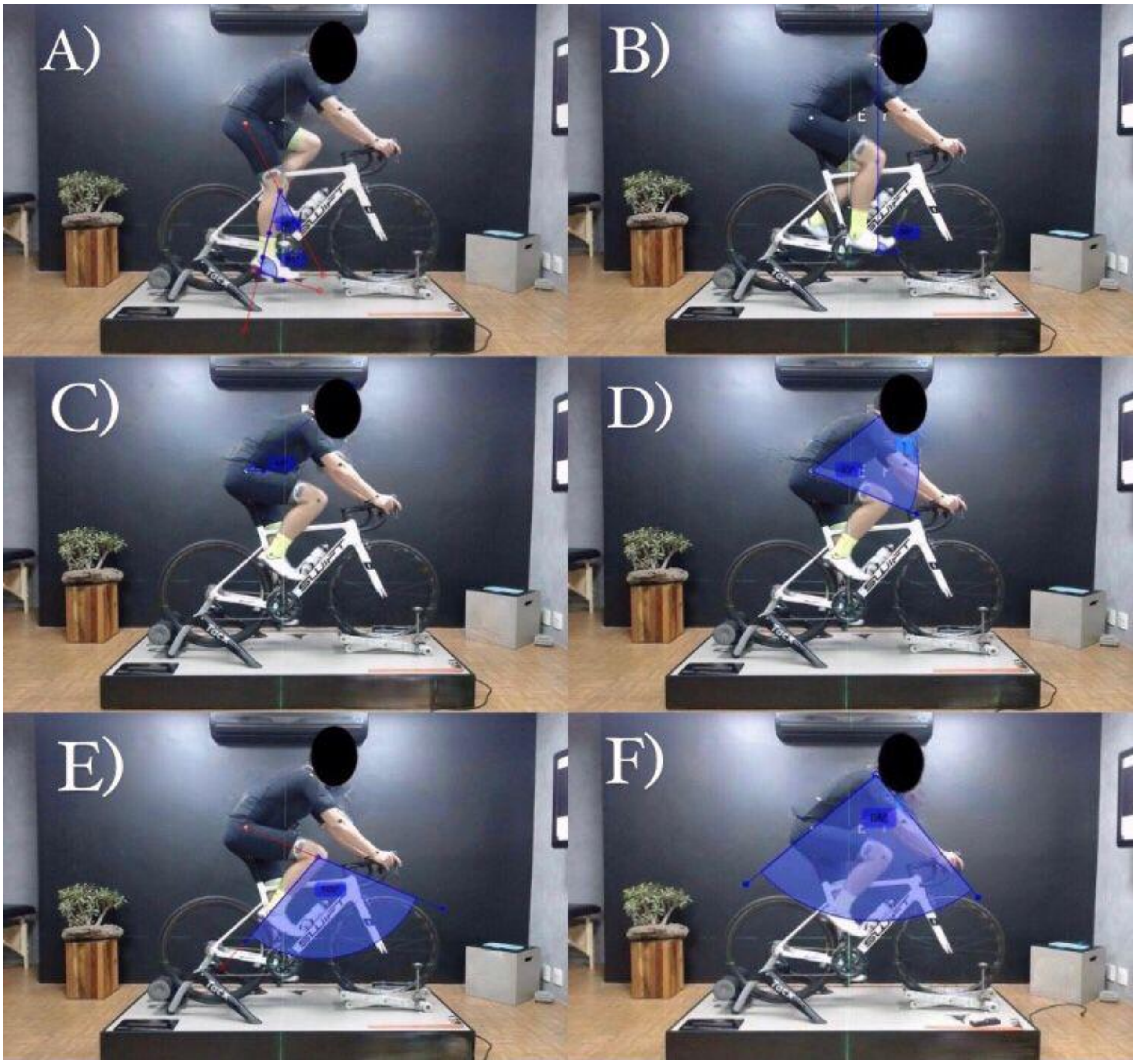

Figure 2. Presentation of the measurement standard to be used by the bike fitters participating in the research for each of the seven measures of interest. A) Ankle plantar flexion angle and knee extension angle; B) KOPS (Knee over pedal spindle); C) Trunk flexion angle; D) Hip flexion angle; E) Knee flexion angle; F) Reach angle (Armpit).

2.4 Statistical Analysis

All statistical analyses were performed in Python (version 3.11), using libraries such as SciPy, Pandas, Matplotlib, and Seaborn. The distribution of each angular variable was assessed using the Shapiro–Wilk test ($\alpha = 0.05$) and Q–Q plots. Variables with approximate normal distribution were summarized as mean ± SD with 95% CI, while non-normal variables were reported as median (IQR) with bias-corrected and accelerated (BCa) bootstrap CIs. Among the seven angular variables, trunk flexion, hip flexion, and knee flexion followed an approximately normal distribution (Shapiro–Wilk $p \geq 0.05$), while ankle plantar flexion, knee extension, reach angle, and KOPS did not ($p < 0.05$).

Variability was quantified using standard deviation (SD), coefficient of variation (CV%), interquartile range (IQR), and tolerance intervals. For the CV%, 95% confidence intervals were estimated using a bias-corrected and accelerated (BCa) bootstrap procedure with 10,000 resamples, given the non-normal distribution of this metric. One-sided chi-square variance tests ($\alpha = 0.05$) were applied against pre-defined tolerance thresholds (2–10% of the mean for each angle). To account for multiple comparisons, p-values were adjusted using the Holm correction.

Because all raters analyzed the same video (single target), traditional approaches such as intraclass correlation coefficients (ICC) or repeated-measures ANOVA were not applicable, as these methods require independent repeated measurements across multiple targets. Instead, agreement was assessed by: (i) bias and bootstrap 95% limits of agreement relative to the sample consensus (median), (ii) robust dispersion comparisons across subgroups of experience level and software use (Brown–Forsythe test), and (iii) Fleiss' κ with 95% CI for categorical KOPS classification. Visualization included box/violin plots and Bland–Altman-style plots to illustrate the distribution of differences relative to consensus, though not as formal reliability tests.

Sensitivity analyses were performed restricting the sample to raters with ≥3 years of clinical experience and those reporting regular Kinovea® use, to examine whether expertise influenced consistency. A two-sided $\alpha = 0.05$ was adopted, with Holm corrections applied where appropriate.

3 Results

The sample consisted of 53 professionals working in the field of bike fitting. The vast majority were male (n = 51; ≅96.39%), with only two female participants (≅3.76%). Participants were distributed across 17 Brazilian states, with the highest representation from São Paulo (n = 17; ≅31.96%), followed by Minas Gerais (n = 10; ≅18.80%), Bahia and Piauí (each with n = 4; ≅7.52%), Espírito Santo (n = 3; ≅5.64%), and Mato Grosso, Rio de Janeiro, and Rio Grande do Sul (each with n = 2; ≅3.76%). One participant (≅1.88%) was reported in each of the following states: Alagoas, Amapá, Ceará, Federal District, Goiás, Pará, Paraíba, Pernambuco, and Santa Catarina.

Regarding the software used in bike fitting studios, the most commonly reported tools were Kinovea (n = 12; ≅22.56%), VelogicFit (n = 11; ≅20.68%), Guru (n = 9; ≅16.92%), and Retül (n = 8; ≅15.04%). Other software included Bike Fast Fit (n = 7; ≅13.16%), STT Systems (n = 2; ≅3.76%), Infinit (n = 1; ≅1.88%), and other unspecified platforms (n = 3; ≅5.64%).

Participants ranged in age from 24 to 59 years, with a mean age of 40.69 ± 7.38 years (95% CI: 38.66–42.73). Professional experience as bike fitters varied from 0.25 to 20 years, with a mean of 6.10 ± 4.61 years (95% CI: 4.83–7.37).

Normality was assessed via the Shapiro–Wilk test and visual inspection of Q–Q plots.

Variables departing from normality ($p < 0.05$): Armpit, Hip Flexion, Knee Extension, Knee Flexion, Plantar Flexion, Trunk Flexion. Table 1 presents the descriptive statistics for all angular variables evaluated by the raters. For each joint angle, we report mean, standard deviation, minimum and maximum values, 95% confidence intervals (CI) for the mean, as well as the coefficient of variation (CV%) with bootstrap-based 95% CIs. Overall, variability was lowest for knee flexion (CV = 2.03%, 95% CI: 1.59–2.67) and highest for knee extension (CV = 6.21%, 95% CI: 4.90–8.68). These findings indicate that, while central tendency values were relatively consistent across evaluators, the degree of dispersion varied depending on the joint angle analyzed.

Table 1. Descriptive statistics for angular variables.

Variable	N	Mean	SD	Min	Max	95% CI (Mean) Lower	95% CI (Mean) Upper	CV (%)	95% CI (CV%) Lower (BCa)	95% CI (CV%) Upper (BCa)
Plantar Flexion	53	90.76	4.39	80.0	101.0	89.55	91.97	4.84	3.73	6.16
Knee Extension	53	46.1	2.86	36.0	52.0	45.31	46.89	6.21	4.9	8.68
Trunk Flexion	53	50.46	2.95	48.6	69.0	49.65	51.27	5.84	2.54	11.9
Hip Flexion	53	72.02	2.68	63.0	80.1	71.28	72.76	3.73	2.71	5.18
Knee Flexion	53	108.33	2.2	102.3	115.0	107.72	108.93	2.03	1.59	2.67
Armpit	53	78.28	2.41	70.0	85.0	77.62	78.95	3.08	2.28	4.23

Descriptive statistics of angular measurements obtained by 53 bike fitters. Data are presented as mean, standard deviation (SD), minimum, maximum, and 95% confidence interval (CI) for the mean. Coefficient of variation (CV%) is shown with bootstrap bias-corrected and accelerated (BCa) 95% confidence intervals.

Figure 3 presents the Q-Q plots for each joint angle assessed in the study. These plots compare the observed data to a theoretical normal distribution, enabling a visual evaluation of normality. Deviations from the reference line indicate potential departures from normality, which were more pronounced in the trunk flexion, hip flexion, and plantar flexion angles.

Figure 4 illustrates the Bland–Altman-style plots for each of the angular variables. Across all measures, the mean bias was close to zero, but the width of the empirical limits of agreement varied substantially. Plantar flexion and trunk flexion showed the widest dispersion, with several raters deviating more than ±10° from the consensus. In contrast, knee flexion and armpit angle presented narrower limits, indicating greater consistency. These findings reinforce the descriptive and inferential results presented in Tables 1 and 3, highlighting that reliability differs depending on the joint angle analyzed.

Figure 5 displays box plots of the angular measurements for each of the joints analyzed. Each plot illustrates the minimum and maximum values, median, quartiles, and potential outliers, serving as a useful tool for identifying trends, dispersion, and anomalies in the data.

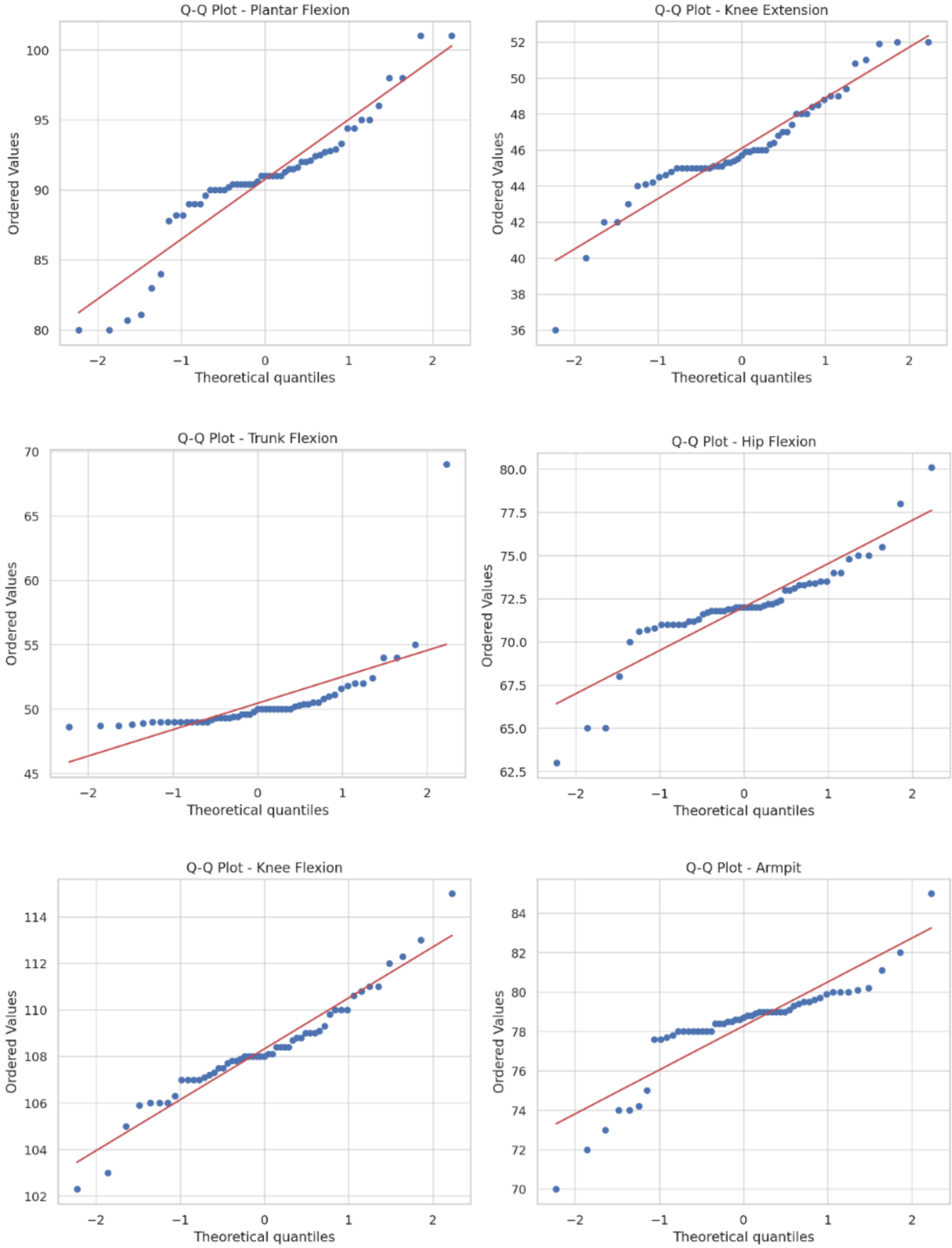

Figure 3. Q-Q plots of angular measurements for each joint angle. The plots compare the theoretical quantiles of a normal distribution (red line) with the ordered values of the data (blue dots). Deviations from the reference line indicate departures from normality. Greater deviations are observed in Trunk Flexion, Hip Flexion, and Plantar Flexion, suggesting non-normal distributions for these variables. Knee Flexion shows the best adherence to normality.

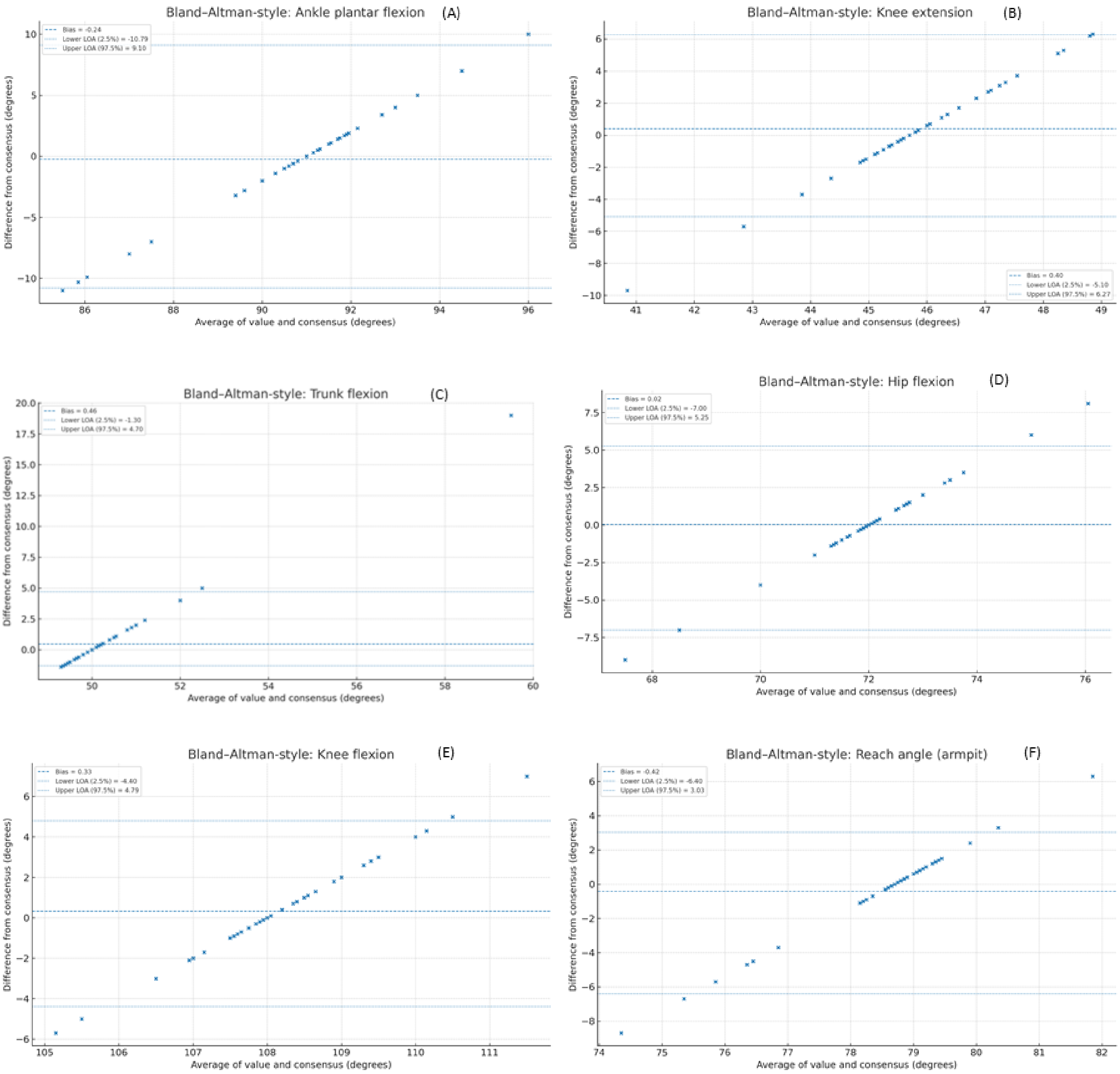

Figure 4. Bland–Altman-style plots for angular variables measured by 53 bike fitters. Differences between individual raters and the consensus (median) are plotted against the average of the two values. The dashed line represents the mean bias, and dotted lines indicate the empirical 95% limits of agreement (2.5th and 97.5th percentiles). Panels: (A) ankle plantar flexion; (B) knee extension; (C) trunk flexion; (D) hip flexion; (E) knee flexion; (F) reach angle (armpit).

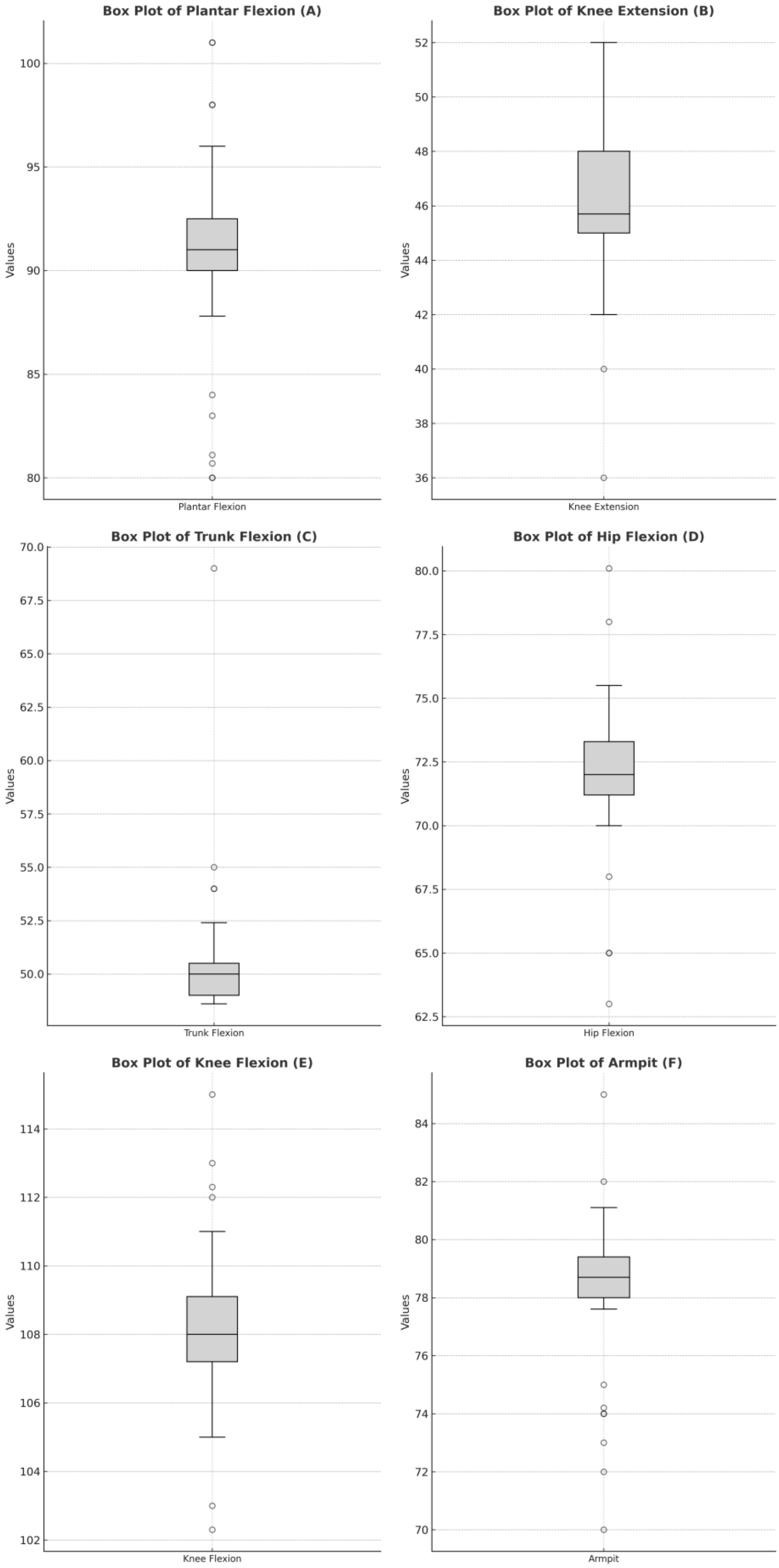

Figure 5. Box plots of the angular measurements obtained for each joint variable. The plots display the distribution, central tendency, and variability of the data, including median, interquartile range, and potential outliers. A - plantar flexion; B - knee extension; C - trunk flexion; D - hip flexion; E - knee flexion; F - armpit region.

Table 2. Chi-square variance tests of joint angle measurements against predefined tolerance thresholds (2–10% of the mean).

Joint Angle	Chi-square 2%	p-value 2%	Chi-square 4%	p-value 4%	Chi-square 6%	p-value 6%	Chi-square 10%	p-value 10%	p_value_holm 2%	p_value_holm 4%	p_value_holm 6%	p_value_holm 10%
Plantar Flexion	304.83	0.0000*	76.21	0.0160*	33.87	0.9757	12.19	1.0000	0.0000	0.0480	1.0000	1.0000
Knee Extension	501.44	0.0000*	125.36	0.0000*	55.72	0.3369	20.06	1.0000	0.0000	0.0000	0.6737	1.0000
Trunk Flexion	443.62	0.0000*	110.91	0.0000*	49.29	0.5811	17.74	1.0000	0.0000	0.0000	1.0000	1.0000
Hip Flexion	180.40	0.0000*	45.10	0.7399	20.04	1.0000	7.22	1.0000	0.0000	1.0000	1.0000	1.0000
Knee Flexion	53.56	0.4140	13.39	1.0000	5.95	1.0000	2.14	1.0000	1.0000	1.0000	1.0000	1.0000
Armpit	123.31	0.0000*	30.83	0.9914	13.70	1.0000	4.93	1.0000	0.0000	1.0000	1.0000	1.0000

Results of chi-square variance tests for angular variability across predefined thresholds (2%, 4%, 6%, and 10% of the mean). For each joint angle, the chi-square statistic, and raw and Holm-adjusted p-values are presented. Values marked with an asterisk (*) indicate statistical significance ($p < 0.05$).

Table 2 summarizes the chi-square variance tests conducted to evaluate whether the variability of each joint angle exceeded predefined tolerance thresholds (2%, 4%, 6%, and 10% of the mean). After applying Holm corrections for multiple comparisons, significant deviations from the strictest thresholds (2% and 4%) were observed for plantar flexion, knee extension, trunk flexion, and hip flexion, whereas knee flexion and armpit angles remained within acceptable limits across all thresholds.

Agreement analyses relative to the group consensus (median) are presented in Table 3. Across all angular variables, the mean bias was close to zero, indicating minimal systematic deviation from the consensus. However, the empirical 95% limits of agreement revealed varying degrees of dispersion, with the widest interval observed for plantar flexion (−10.79 to 9.10°) and hip flexion (−7.0 to 5.25°). In contrast, knee flexion and trunk flexion exhibited narrower limits of agreement, suggesting greater consistency among raters for these measures.

To evaluate differences in measurement variability across subgroups, Brown–Forsythe tests of dispersion were performed (Table 4). No statistically significant differences were found when comparing bike fitters with less than 3 years of experience to those with 3 or more (all p > 0.05), although a non-significant trend toward greater variability was observed for knee flexion (p = 0.0838). Similarly, no significant differences emerged across subgroups defined by the software used (Kinovea, VelogicFit, Guru, Retül, Bike Fast Fit, or others), with all p-values > 0.24. These results suggest that neither professional experience nor software choice meaningfully influenced the consistency of angular measurements in this sample.

Table 3. Agreement relative to group consensus (median): bias and empirical 95% limits of agreement.

Variable	Bias (mean diff vs median)	Empirical LOA Lower (2.5%)	Empirical LOA Upper (97.5%)
Plantar Flexion	-0.242	-10.79	9.1
Knee Extension	0.398	-5.1	6.27
Trunk Flexion	0.462	-1.3	4.7
Hip Flexion	0.023	-7.0	5.25
Knee Flexion	0.326	-4.4	4.79
Armpit	-0.419	-6.4	3.03

Agreement of raters' measurements relative to the group consensus (median). Bias represents the mean difference between individual ratings and the consensus. Empirical 95% limits of agreement (LOA) were derived using bootstrap resampling (2.5th and 97.5th percentiles).

Table 4. Brown–Forsythe tests of dispersion across subgroups.

Variable	Brown–Forsythe Statistic	p_value
Experience (<3 years vs. ≥3 years)		
Plantar Flexion	0.5581	0.4584
Knee Extension	0.3687	0.5464
Trunk Flexion	0.0072	0.9325
Hip Flexion	0.0789	0.7800
Knee Flexion	3.1102	0.0838
Armpit	0.5622	0.4568
Software (Kinovea, VelogicFit, Guru, Retül, Bike Fast Fit, and others)		
Plantar Flexion	0.3229	0.8964
Knee Extension	0.6411	0.6696
Trunk Flexion	0.4484	0.8121
Hip Flexion	1.3835	0.2494
Knee Flexion	0.9092	0.4841
Armpit	0.7196	0.6123

Brown–Forsythe tests of homogeneity of dispersion across subgroups. Factors tested included professional experience (<3 years vs. ≥3 years) and software used for bike fitting (Kinovea, VelogicFit, Guru, Retül, Bike Fast Fit, and others). No statistically significant differences in variability were identified (p > 0.05).

Regarding KOPS positioning, the only categorical variable included in the study, 49.1% of participants (n = 26) reported using a neutral positioning, while 30.2% (n = 16) reported a negative positioning and 20.8% (n = 11) adopted a positive positioning. KOPS classifications showed 49.1% agreement with the modal category (Neutro; 95% CI: 36.1–62.1%; N = 53). Given the single-target design, Fleiss' κ was not estimable; the modal agreement proportion with Wilson 95% CI is reported as an alternative.

4 Discussion

The findings of this study reveal a high degree of inter-examiner variability in angular measurements performed using Kinovea® software, even when all evaluators analyzed the same video under standardized conditions. For more stringent angular variation thresholds (2% and 4%), several variables—including knee extension, trunk flexion, and plantar flexion—demonstrated significantly greater variance than expected, as confirmed by chi-square hypothesis testing. These results indicate that, despite the provision of visual measurement standards and detailed instructions, inconsistencies persist in how bike fitters implement 2D kinematic analysis protocols.

The coefficients of variation (CV%) for most joint angles exceeded 4%, and in some cases surpassed 6%, suggesting a moderate to high degree of inconsistency among evaluators. The presence of multiple outliers and asymmetries in the box plots, alongside rejection of normality in several variables, supports the interpretation that this variation is not merely random but may stem from methodological inconsistencies, differing levels of professional experience, or limited proficiency with the software interface.

Such findings highlight serious concerns about the reproducibility and clinical reliability of 2D motion analysis in bike fitting, particularly when conducted by different professionals without unified training. In the absence of standardized procedures and ongoing education, the likelihood of inaccurate measurements increases, which may result in suboptimal—or potentially harmful—adjustments to the cyclist–bike configuration. This underscores the importance of developing robust evaluation protocols and, where feasible, adopting semi-automated or 3D technologies to enhance the reliability of professional bike fitting practices.

This study examined the consistency of angular measurements taken by bike fitters from various regions of Brazil using Kinovea®. As cycling continues to gain popularity, so does the demand for services like bike fitting (Lobo et al., 2018; Robidoux, 2021). Effective adjustments to the cyclist–bike system (CBS) require the integration of anamnesis, physical assessment, and a reliable kinematic analysis protocol. Standardized evaluation procedures are essential to reduce systematic biases that can affect decision-making and compromise cyclists' health and performance (Quesada et al., 2019; Scoz et al., 2022).

A lack of attention to the ergonomic aspects of the CBS may contribute to reports of pain, discomfort, and injury among cyclists—factors that can reduce motivation or hinder participation (Scoz et al., 2022). Proper adjustments to components and footwear are known to reduce injury risk (Quesada et al., 2019; Scoz et al., 2022). Even minor corrections can significantly improve comfort and performance, whereas poor or absent adjustments can lead to discomfort, microtrauma, or serious musculoskeletal issues (Scoz et al., 2021).

The study's sample included 53 qualified Brazilian professionals. According to the International Bicycle Fitting Institute (2019), no formal degree is required to practice as a bike fitter. It is plausible that this variability in educational background and training contributed to the observed inconsistencies in measurement.

Experience length may also influence measurement precision. However, with an average of six years of practice, the participants were not novices. Thus, lack of experience is unlikely to be the primary cause of the observed variation.

The data from this study show that the most commonly used software in the participants' daily routines were KINOVEA® and VelogicFit. The former is a free tool for 2D analysis (Norris et al., 2011; Thewlis et al., 2013; Puig-Diví et al., 2019), while the latter provides 3D analysis and requires a paid license and monthly subscription (Velogicfit, 2023). In this study, all bike fitters used KINOVEA® to perform the 2D kinematic analysis. Importantly, although all participants had previous experience with Kinovea®, many no longer use it regularly. This diminished familiarity may have hindered their ability to apply measurement techniques with precision. Like any tool, Kinovea® requires consistent use and interface fluency to ensure measurement reliability. KINOVEA® has been recommended for use in three primary contexts: sports (Sañudo et al., 2016), research (Guzmán-Valdivia et al., 2013), and the validation of new technologies (Padulo et al., 2015). Like any measurement instrument, however, it requires training and familiarity with its interface to ensure reliable data.

Regarding the joint angles of interest, there was considerable variability in the measurements obtained by different bike fitters. The greatest discrepancies were observed in the knee extension and trunk flexion angles. Depending on the value recorded during kinematic analysis, bike fitters may arrive at opposing conclusions—for instance, one might decide to raise the saddle, while another might choose to lower it. Such divergent decisions could negatively impact the cyclist's positioning, potentially compromising health, performance, and perceived comfort (Bini & Quesada, 2022).

This finding is particularly noteworthy given that all participants analyzed the same video of the same cyclist, with anatomical markers pre-positioned by the researchers. The analysis was performed using widely adopted software in the field, and participants were provided with a detailed set of reference images illustrating how each angle should be measured. Therefore, the high degree of variability observed among evaluators was unexpected.

Several factors may partially explain the high variability observed. Some have already been mentioned, such as the bike fitters' academic training or limited familiarity with the motion analysis software. Another important source of variability relates to the method of angle measurement, specifically whether raters used the real or the supplementary angle. In routine practice, some fitters measure the real angle, defined by the intersection of lines connecting anatomical landmarks, while others opt for the supplementary angle, two values whose sum equals 180° (Silva, 2023). In this study, the use of the supplementary angle for knee extension and flexion was both recommended and illustrated; however, fitters unfamiliar with this convention may have encountered difficulties in measurement and interpretation. If raters did not apply the same convention consistently, part of the observed variability may reflect methodological inconsistency rather than true differences in assessment. This potential source of bias underscores the importance of establishing explicit and standardized guidelines for angle definition in 2D motion analysis to improve reproducibility and inter-rater reliability.

Additionally, a 40-second video contains approximately 600 frames, making it unlikely that all fitters selected the exact same frame for angle analysis. However, it is believed that frame selection alone does not account for the degree of variability observed. If this were the case, the validity of the bike fit process itself would be undermined, as it would imply that each frame represents a distinct body position capable of influencing clinical decisions.

Therefore, it is likely that the measurement procedure itself is the primary contributor to the observed variability. During this process, the bike fitter must draw lines between markers. Each marker used in the study was 20 mm in diameter, and depending on the exact point where the fitter begins the line, the resulting angle may be slightly more or less inclined. Furthermore, once the lines are drawn, the angle must be projected onto them. In the software, even small adjustments to the vertex of the angle at the line intersection can result in noticeable changes in the measurement (Balsalobre et al., 2014). To reduce this issue, it is essential to standardize the measurement protocol in order to improve both intra- and inter-examiner reliability. Although reliability was not evaluated in the present study, this aspect will be addressed in future research by the study group.

The only qualitative variable in the study was KOPS (knee over pedal spindle). For this measure, bike fitters were instructed to pause the video with the crank arm positioned parallel to the ground and draw a perpendicular line from the pedal spindle upward to at least the level of the marker placed on the lateral femoral condyle. According to the protocol, if the marker was located behind the line, KOPS was classified as negative; if directly on the line, neutral; and if in front of the line, positive (Burt, 2022). Despite its qualitative nature, this variable also demonstrated significant variability among fitters. This may be explained by procedural errors, particularly in selecting a frame where the crank arm was truly parallel to the ground. Another source of error could be improper line drawing—if the line is not exactly 90° relative to the ground or is misaligned with the pedal spindle, the KOPS classification becomes unreliable.

The absence of significant differences in measurement variability across subgroups defined by professional experience and software choice has relevant practical implications. First, it suggests that inter-examiner variability is not simply a function of years of practice, indicating that even experienced fitters may be subject to inconsistency when analyzing 2D kinematics. This highlights the need for structured training protocols and clearer operational definitions for anatomical landmarking, rather than relying solely on accumulated clinical experience. Second, the lack of variability across different motion analysis tools reinforces that the primary source of inconsistency is not the software itself but the examiner's methodological approach. Together, these findings emphasize the importance of developing standardized guidelines for video analysis in bike fitting, which may improve reproducibility and reduce subjectivity in clinical and performance contexts.

The variability observed in this study should not be interpreted as a limitation of the Kinovea® software. Multiple investigations have confirmed its validity and reliability in sports and clinical contexts. For instance, Kinovea® has been used to detect improvements in pelvic drop and frontal plane knee angles in runners with medial tibial stress syndrome. Harrington et al. (2023) reported good-to-excellent intra- and inter-rater reliability (ICC = 0.80–0.98) with acceptable error (<5°) during remote squat and lunge assessments. Vicente-Pina et al. (2025) found excellent agreement with a 3D motion capture system for pelvic motion (ICC > 0.90), and Pueo et al. (2020) validated the Smartphone-Kinovea method for vertical jumps, showing almost perfect agreement with laboratory systems (ICC > 0.98). Together, these findings indicate that Kinovea® is a robust and reliable tool, and that the variability in our study is more likely due to differences in evaluators' methods, such as frame selection, anatomical landmarking, and prior experience, than to limitations of the software itself.

In addition to continuous training and professional development, which may reduce measurement variability, the use of automated systems with marker tracking is also recommended. These systems employ motion sensors, optical markers, and advanced equipment and software to eliminate the need for manual angle measurement by the bike fitter (Eltoukhy et al., 2012; Gonzáles et al., 2020). However, such systems are costly, require specialized maintenance, and demand specific training (Velogicfit, 2023). These factors increase the operational costs of bike fitting services, which are often passed on to the client / cyclist. Depending on the final cost, some cyclists may opt not to undergo a bike fit. Thus, it is essential to find a balance: ensuring that cyclists have access to reliable assessments that genuinely inform professional decision-making, while maintaining affordability and accessibility.

This study has some limitations that should be acknowledged. First, only a single video recording was analyzed, which restricts the generalizability of the findings. Additionally, intra-rater reliability was not assessed, which limits direct comparison with other reliability studies. Although more robust reliability metrics such as ICC or repeated-measures ANOVA are often recommended, these approaches were not applicable to our single-target, inter-examiner design. Instead, alternative agreement analyses (bias, bootstrap LOA, Brown–Forsythe, Fleiss' κ) were employed. Finally, the sample, although relatively large compared to previous research, was obtained by convenience sampling, which may introduce bias in terms of participants' representativeness.

5 Practical Applications

- 2D kinematic analysis with software like Kinovea® revealed high inter-examiner variability, even when all fitters analyzed the same cyclist under identical conditions.
- Inconsistent angular measurements during bike fitting may lead to inappropriate adjustments, with potential negative effects on comfort, performance, and injury risk.
- Seemingly small methodological choices—such as video frame selection or anatomical landmark placement—can generate meaningful differences in the outcomes of a bike fit.
- Standardized protocols and structured training are critical to enhance measurement reliability, reduce examiner subjectivity, and ensure safer, more effective fitting decisions.
- Cyclists, coaches, and clinicians should be aware that manual 2D assessments may vary across professionals, reinforcing the importance of working with experienced or well-trained fitters.

6 Conclusions

The study revealed substantial inter-examiner variability in 2D kinematic measurements using Kinovea® software, particularly for knee extension, trunk flexion, and plantar flexion angles, with coefficients of variation exceeding 5% in several cases. These findings indicate that, even under standardized conditions, the reliability of manual 2D measurements can be compromised when performed by different professionals. To enhance consistency and clinical utility, it is essential to adopt standardized measurement protocols and promote continuous examiner training. Future research should explore strategies to reduce variability, including semi-automated tools and structured qualification programs for bike fit practitioners.

Acknowledgments: The authors would like to thank the Centro Universitário de Formiga (UNIFOR-MG) and the National Council for Scientific and Technological Development (CNPq) for their support of this research.

Conflicts of Interest: The authors APP, GCV, and MRM are directly involved in Bike Fit practice. The authors declare no other conflicts of interest.

References

1. Robidoux, C. G. (2022). A practical approach to the evaluation of a cyclist with overuse injury. Physical Medicine and Rehabilitation Clinics of North America, 33(1), 15–29. doi: 10.1016/j.pmr.2021.08.002
2. Kotler, D. H., Babu, A. N., & Robidoux, G. (2016). Prevention, evaluation, and rehabilitation of cycling-related injury. Current Sports Medicine Reports, 15(3), 199–206. doi: 10.1249/JSR.0000000000000262
3. Scoz, R. D., Espindola, T. R., Santiago, M. F., de Oliveira, P. R., Alves, B. M. O., Ferreira, L. M. A., & Amorim, C. F. (2022). Long-term effects of a kinematic bikefitting method on pain, comfort, and fatigue: A prospective cohort study. International Journal of Environmental Research and Public Health, 19(19), 12949. doi: 10.3390/ijerph191912949
4. Bini, R., Daly, L., & Kingsley, M. (2019). Changes in body position on the bike during seated sprint cycling: Applications to bike fitting. European Journal of Sport Science, 19(6), 842–849. doi: 10.1080/17461391.2019.1593843
5. Puig-Diví, A., Escalona-Marfil, C., Padullés-Riu, J. M., Busquets, A., Padullés-Chando, X., & Marcos-Ruiz, D. (2019). Validity and reliability of the Kinovea program in obtaining angles and distances using coordinates in 4 perspectives. PLOS ONE, 14(6), e0216448. doi: 10.1371/journal.pone.0216448
6. Scoz, R. D., Espindola, T. R., Santiago, M. F., Oliveira, P. R., Alves, B. M. O., Ferreira, L. M. A., & Amorim, C. F. (2021). Validation of a 3D camera system for cycling analysis. Sensors, 21(13), 4473. doi: 10.3390/s21134473
7. Bini, R., & Hume, P. (2020). Confiabilidade dentro e entre as sessões de um sistema de força de pedal para saída de potência e medições de eficácia de força de pedal \[Article in Portuguese]. Human Movement, 21, 69–78. doi: 10.5114/hm.2020.94188
8. Lobo, Z., Andrade, V., Rodrigues, J., Marino, F., & Binatti, G. (2018). Perfil do ciclista (Relatório técnico). Transporte Ativo & LABMOB-UFRJ. https://www.labmob.ufrj.br/relatorio-perfil-do-ciclista-2018
9. Quesada, J. I. P., Kerr, Z. Y., Bertucci, W. M., & Carpes, F. P. (2019). The association of bike fitting with injury, comfort, and pain during cycling: An international retrospective survey. European Journal of Sport Science, 19(6), 842–849. doi: 10.1080/17461391.2019.1593843
10. International Bicycle Fitting Institute. (2019). Certificação profissional de bike fit. https://www.bikefitinstitute.com
11. Norris, B. S., & Olson, S. L. (2011). Concurrent validity and reliability of a two-dimensional video analysis for hip and knee joint motion during mechanical lifting. Physiotherapy Theory and Practice, 27(7), 521–530. doi: 10.3109/09593985.2010.509889
12. Thewlis, D., Bishop, C., Daniell, N., & Paul, G. (2013). Can low-cost, next-generation motion capture systems provide comparable spatial accuracy to high-end systems? Journal of Applied Biomechanics, 29(1), 112–117. doi: 10.1123/jab.29.1.112
13. Velogic Studio. (2023). Efficient fits, happy customers. https://www.velogicfit.com
14. Sañudo, B., Rueda, D., Pozo-Cruz, B. D., de Hoyo, M., & Carrasco, L. (2016). Validity of a video-based software analysis package for quantifying movement velocity in resistance exercises. Journal of Strength and Conditioning Research, 30(10), 2934–2939. doi: 10.1519/JSC.0000000000001390

15. Guzmán-Valdivia, C. H., Blanco-Ortega, A., Oliver-Salazar, M. A., & Carrera-Escobedo, J. L. (2013). Analysis of therapeutic movement of lower limbs using Kinovea. International Journal of Computer Science and Engineering, 3(2), 359–365.
16. Padulo, J., Vando, S., Chamari, K., Chaouachi, A., Bagno, D., & Pizzolato, F. (2015). Validity of the MarkWiiR for kinematic analysis during walking and running gait. Biology of Sport, 32(1), 53–58. doi: 10.5604/20831862.1127288
17. Bini, R., & Priego-Quesada, J. (2022). Methods to determine saddle height in cycling and implications of changes in saddle height on performance and injury risk: A systematic review. Journal of Sports Sciences, 40(4), 386–400. doi: 10.1080/02640414.2021.1994727
18. Silva, M. N. P. (n.d.). Ângulos complementares, suplementares e adjacentes. Brasil Escola. Retrieved October 17, 2023, from https://brasilescola.uol.com.br/matematica/angulos-complementares-angulos-suplementares-angulos-.htm
19. Balsalobre-Fernandez, C., Tejero-Gonzalez, C. M., del Campo-Vecino, J., & Bavaresco, N. (2014). Simultaneous validity and reliability of a low-cost, high-speed camera-based method for measuring vertical jump flight time. Journal of Strength and Conditioning Research, 28(2), 528–533. doi: 10.1519/JSC.0b013e31829b551c
20. Burt, P. (2014). Bike fit: Optimise your bike position for high performance and injury avoidance. Bloomsbury Sport.
21. Eltoukhy, M., Asfour, S., Thompson, C., & Latta, L. (2012). Evaluation of the performance of digital video analysis for human motion tracking: Dartfish tracking system. International Journal of Scientific and Engineering Research, 3(1), 1–6.
22. Gonzales, P., Koutsou, A., Gómez, A., Tejada, M., Page, J., & Rueda, F. (2020). Reliability of Kinovea® software and agreement with a three-dimensional motion system for gait analysis in healthy subjects. Journal of Biomechanics, 103, 109715. doi: 10.1016/j.jbiomech.2020.109715
23. Harrington MS, Adeyinka IC, Burkhart TA. Intra- and interrater reliability and agreement of a method to quantify lower-extremity kinematics using remote data collection. J Sport Rehabil. 2023;32(7):690-696. doi: 10.1123/jsr.2022-0453
24. Pueo B, Penichet-Tomas A, Jiménez-Olmedo JM. Validity, reliability and usefulness of smartphone and Kinovea motion analysis software for direct measurement of vertical jump height. Physiol Behav. 2020;227:113144. doi: 10.1016/j.physbeh.2020.113144
25. Vicente-Pina L, Sánchez-Rodríguez R, Ferrández-Laliena L, Heredia-Jimenez J, Müller-Thyssen-Uriarte J, Monti-Ballano S, Hidalgo-García C, Tricás-Moreno JM, Lucha-López MO. Validity and reliability of Kinovea® for pelvic kinematic measurement in standing and sitting positions with 45° hip flexion. Sensors (Basel). 2025;25(1):250. doi: 10.3390/s25010250