theoretical and applied ergonomics

MDPI

Article

Development and Validation of the Manual Integrated Ergonomics Risk Assessment Tool (MIERAT): Field Application in a Small-Scale Automotive Service Environment

Khumbuzile Mdlalose * and Peter Madindwa Mashinini

Department of Mechanical and Industrial Engineering, Faculty of Engineering and the Built Environment, University of Johannesburg, Johannesburg 2006, South Africa; mmashinini@uj.ac.za
* Correspondence: khumbuzilen@uj.ac.za; Tel.: +27-65-821-2766

Abstract

Work-related musculoskeletal disorders (WRMSDs) remain a major occupational health challenge, and automotive service work is a high-risk context involving awkward postures, repetitive tasks, and sustained static loading. Conventional ergonomic tools capture static posture snapshots and do not reflect how musculoskeletal stress accumulates over time. This study developed the Manual Integrated Ergonomics Risk Assessment Tool (MIERAT), a low-cost, observation-based framework for resource-constrained automotive settings, and conducted an initial field validation. MIERAT combines REBA-based posture severity (PS), exposure duration (D), and a calibrated Time Modifier (TM; 1.0–2.5) into a Cumulative Risk Index (CRI = PS × D × TM), grounded in biomechanical loading theory and aligned with ISO 11226, ISO 11228-1, and ISO 45001. Expert panel review yielded an S-CVI/Ave of 0.95, indicating strong content validity. Duration–CRI associations were very strong for under-vehicle servicing ($\rho = 0.910$), tyre fitment ($\rho = 0.961$), and brake/suspension repair ($\rho = 0.945$), all $p < 0.001$. Because duration enters the CRI equation directly and also determines the TM, these correlations are mathematically expected; they show that the index behaves as designed but are not independent evidence of construct validity. Inter-rater reliability across three observers was excellent (ICC = 0.975–0.997), with moderate-to-perfect categorical agreement (weighted $\kappa = 0.557$–1.000). Task-level profiling identified tyre demounting, wheel alignment, and drain plug loosening as dominant ergonomic hotspots. These findings provide preliminary support for MIERAT as a content-valid and reliable tool for profiling cumulative postural exposure and planning ergonomic interventions in South African automotive workshops. Validation against independent outcomes, such as musculoskeletal symptoms, fatigue, biomechanical loading, or injury data, is needed to establish its construct and predictive validity.

Keywords: ergonomic risk assessment; MIERAT; cumulative risk index; REBA; work-related musculoskeletal disorders; automotive service workers

Check for updates

Academic Editor: Philippe Gorce

Received: 6 July 2026
Revised: 6 October 2026
Accepted: 10 October 2026
Published: 23 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

Work-related musculoskeletal disorders (WRMSDs) are among the most significant occupational health challenges worldwide, affecting an estimated 1.71 billion individuals worldwide and accounting for more than 160 million workplace injuries annually [1,2].

In automotive service environments, particularly small- and medium-scale workshops, workers face an elevated ergonomic burden owing to the manual, non-repetitive,

and physically demanding nature of tasks such as tyre fitment, under-vehicle servicing, and brake and suspension repair. These operations typically involve prolonged trunk flexion, constrained kneeling, sustained overhead reaching, and repeated forceful exertions, collectively generating cumulative musculoskeletal stress that predisposes workers to chronic injury [3–5]. In Africa, automotive repair workers are a particularly vulnerable occupational group. A study of 344 vehicle repair workers in Ethiopia reported a 12-month WRMSD prevalence of 47.7%, with the lower back being the most commonly affected region [3]. Small-scale workshops typically lack the resources and technical infrastructure to adopt advanced ergonomic monitoring technologies or to comply with the formal risk assessment requirements of the Occupational Health and Safety Act 85 of 1993 [6] and the Ergonomics Regulations of 2019 [7,8]. A practical, affordable, and validated ergonomic assessment tool suited to these operational conditions has so far been absent from the literature. Conventional ergonomic assessment tools, including REBA [9], RULA [10], OWAS [11], QEC [12], and OCRA [13], provide valuable information on posture risk but share a fundamental limitation: they deliver static snapshots of biomechanical exposure without integrating the temporal dimension of ergonomic risk. A posture held for 8 s and the same posture sustained for 180 s carry markedly different injury implications, yet existing tools offer no simple, field-applicable means of distinguishing between them [14,15]. Table 1 summarises these limitations and compares the established tools with the proposed MIERAT.

Table 1. Comparative summary of established ergonomic assessment tools and MIERAT: coverage, time integration, limitations, and suitability for small automotive workshops.

Tool	Coverage	Time Integration	Limitations	Small Workshop Suitability
REBA	Whole-body postures; static snapshot	None	Fails to capture temporal accumulation	Moderate
RULA	Upper limbs, neck, trunk	None	Limited for full-body automotive tasks	Low–Moderate
OWAS	General postures; coarse categories	None	Low sensitivity; insufficient detail	Moderate
QEC	Whole-body + worker perception	Indirect (qualitative): subjective repetition/ frequency rating	Subjective; less reliable across raters	Low
OCRA	Repetitive upper-limb actions	Explicit: frequency and recovery-time ratio (not posture-duration-based)	Complex and time-consuming	Low
MIERAT	Whole-body, dynamic posture-time integration	Explicit: posture-duration-based TM scale, 1.0–2.5	Observer training required; novel tool	High (manual, cost-effective, ISO-aligned)

Note: 'Indirect' denotes a qualitative, self-reported acknowledgment of repetition/frequency; 'Explicit' denotes a quantified multiplier or ratio applied to the risk score. MIERAT differs from OCRA in that it weights the duration of individual posture windows rather than shift-level action frequency and recovery ratios.

This paper addresses this theoretical and methodological gap by presenting the development and initial validation of MIERAT. The tool integrates posture severity assessment with explicit temporal exposure modelling through a calibrated Time Modifier (TM) scale, producing a Cumulative Risk Index (CRI) that quantifies the ergonomic burden of each posture window of a task cycle. MIERAT was designed for use in resource-constrained

South African automotive workshops without specialised equipment, advanced ergonomic training, or digital infrastructure.

The overarching aim of this study was to develop MIERAT as a generalisable posture-time ergonomic risk assessment tool and to provide initial evidence of its validity and reliability, using its implementation in a small-scale automotive service workshop as the first empirical test case. Specifically, the objectives of this paper are to: (1) describe the theoretical foundations and design principles of MIERAT; (2) present the tool's conceptual architecture, scoring logic, and process flow; (3) report the content validity, internal consistency of the CRI construct, inter-rater reliability, and task-level CRI profiling results obtained from fieldwork conducted at the Mabopane Automotive Centre (MAC), South Africa; and (4) discuss the tool's theoretical contributions, its practical implications for ergonomic risk management in small automotive workshops, and the further validation required against independent outcomes.

2. Theoretical Foundations

2.1. Biomechanical Loading Theory

Biomechanical loading theory posits that musculoskeletal tissues experience strain proportional to posture configuration, applied load, and biomechanical leverage; injury risk increases with the magnitude and frequency of this strain [16]. In automotive service tasks, trunk flexion, neck flexion and rotation, and constrained knee and hip postures can substantially increase compressive and shear forces at the spinal and peripheral joints, and these forces may exceed tissue tolerance limits when postures are sustained without adequate recovery. MIERAT operationalises this principle through REBA-based body-region scoring, which serves as an established observational proxy for instantaneous biomechanical loading.

2.2. Cumulative Fatigue–Failure Theory

Fatigue–failure theory explains how sub-maximal loads applied repeatedly or continuously cause microscopic tissue damage that accumulates until failure thresholds are reached, resulting in musculoskeletal injury even when individual loading episodes remain below acute injury thresholds [17]. This is directly relevant to automotive service workers, who often sustain postures of low-to-moderate severity over extended task cycles. MIERAT incorporates a Time Modifier to scale posture severity with exposure duration, acknowledging that identical postures can have fundamentally different injury implications depending on how long they are sustained.

2.3. Exposure Accumulation Science

From an exposure science perspective, ergonomic risk is conceptualised as a function of three interacting dimensions—intensity, duration, and frequency [15]:

$$\text{Exposure Risk} = \text{Intensity} \times \text{Duration} \times \text{Frequency} \quad (1)$$

MIERAT translates this tripartite model into a practical scoring system by combining posture severity (intensity) and exposure duration through the CRI formula. Exposure variation models [18] further recognise that risk is shaped by the distribution of postures over time; MIERAT's posture-window methodology captures this variation by treating each stable body position as a discrete, timed episode.

2.4. Systems-Based Occupational Health Theory and ISO Alignment

Systems-based ergonomic theory [19,20] emphasises that ergonomic assessment should be embedded within an organisational safety management cycle rather than con-

ducted as an isolated activity. MIERAT aligns with the Plan–Do–Check–Act (PDCA) framework of ISO 45001 [19], so that risk outputs can feed into structured continuous improvement processes. The tool's posture threshold calibration draws on ISO 11226 [21] (postural load assessment) angle-band recommendations and duration thresholds, ISO 11228-1 [22] (manual handling) for biomechanical strain principles, and ISO/TS 20646 [23] for general guidance on optimising musculoskeletal workload, applied here to automotive service tasks.

3. Tool Design and Architecture

3.1. Core Constructs and the CRI Formula

MIERAT is built on five interrelated constructs that together define how ergonomic risk is measured, quantified, and interpreted:

Posture Severity (PS): A dimensionless score (range 1–15) derived from REBA-based body-region evaluation [9], quantifying instantaneous biomechanical stress.

Exposure Duration (D): The time, in seconds, for which a worker maintains a specific posture window, bounded by observable posture transitions.

Time Modifier (TM): A calibrated dimensionless multiplier (range 1.0–2.5). The duration-band boundaries and multiplier values are not read directly from a single ISO clause; rather, they combine (i) the qualitative endurance/fatigue-accumulation principle underlying ISO 11226's [21] maximum-acceptable-holding-time (MAHT) framework, (ii) high-frequency exposure criteria from ISO 11228-3 [24] and continuous-exposure guidance from ISO/TR 12295 [25], and (iii) cumulative-load and fatigue-onset evidence from Gallagher and Schall [26], McAtamney and Corlett [10], Hedge [27], and Schall, Sesek and Cavuoto [28]. Table 2 presents the full TM scale, and Table 3 gives the derivation basis for each duration band.

Table 2. Time Modifier (TM) scale for MIERAT with ISO 11226-aligned duration bands, category labels, and ergonomic rationale (see Table 3 for the derivation and calibration basis of each duration band and TM value).

Duration Band (s)	Category	TM Value	Ergonomic Rationale
0–10	Very Short	1.0	[A] Minimal exposure; below ISO static posture penalty threshold
11–20	Short	1.2	[A] Slight static load increase; short fatigue accumulation onset
21–30	Moderate	1.5	[A] Approaching fatigue threshold for sustained static postures
31–60	Long	1.8	[A] Significant physiological strain; ISO 11226 MAHT curves indicate risk elevation
61–90	Prolonged	2.0	[A] High static exposure; substantial fatigue risk per ISO/TR 12295 [25]
91–120	Very Prolonged	2.2	[A] Continuous muscle loading likely; vascular and nerve compression risk
>120	Extended	2.5	[A] Very high cumulative loading; ISO 11226 endurance data indicate unacceptable hold for most postures

[A] denotes a duration boundary adapted from the qualitative principle in the cited standard(s) rather than copied verbatim as a numeric threshold; the derivation and calibration basis of each band is given in Table 3.

Table 3. Derivation and calibration of the Time Modifier scale.

Duration Band → TM	Basis
≤10 s → TM = 1.0 [A]	ISO/IEA static-posture definition (static load typically defined as holding >4 s); ISO 11226 MAHT tables only begin penalising after a few seconds. A ≤10 s hold is at or below this trigger and within the acceptable range for most neutral/moderate postures, so no additional penalty is applied.
11–30 s → TM = 1.2–1.5 [A]	ISO 11228-3/OCRA [24] high-frequency framing treats short cycle times (≤~30 s) as elevating exposure when actions recur above ~30/min. Repeated short-cycle postures in this range accumulate exposure faster than isolated brief holds, justifying a step above 1.0.
31–60 s → TM = 1.8 [A]	ISO 11226 [21] MAHT curves indicate that holding times of tens of seconds can already exceed acceptable limits at non-neutral joint angles; ISO/TR 12295 [25] reduces the allowable frequency as continuous duration extends.
61–120 s → TM = 2.0–2.2 [A]	ISO/TR 12295 [25] identifies continuous blocks approaching one minute or more as a key risk boundary; ISO 11226 [21] endurance guidance implies MAHT is breached for many postures within this range.
>120 s → TM = 2.5 [A]	ISO 11226 [21] Annex B endurance data show acceptable hold times shrink sharply with joint angle beyond two minutes; ISO 11228-1/-3 de-rating logic likewise penalises prolonged continuous duration. The multiplier is capped at 2.5 to remain within the ISO tradition of steep de-rating for prolonged exposure.

[A] denotes a duration boundary adapted from the qualitative principle in the cited standard(s) rather than copied verbatim as a numeric threshold. The specific multiplier values (1.0–2.5) were set through expert-panel calibration (see the expert panel review in Section 5.1), informed by the fatigue-accumulation literature of Gallagher and Schall [26], McAtamney and Corlett [10], Hedge [27], and Schall, Sesek and Cavuoto [28].

To make the derivation transparent, Table 3 maps each duration band and multiplier step in Table 2 to its standards-based and empirical justification, and distinguishes the band boundaries adapted from ISO principles from the multiplier values calibrated by the expert panel.

Cumulative Risk Index (CRI): The primary ergonomic output metric, integrating posture severity and temporal exposure into a single score expressed in severity-seconds (adapted from [15,29]):

$$CRI = PS \times D \times TM \quad (2)$$

where PS = posture severity score (REBA-derived, 1–15), D = exposure duration (seconds), and TM = time modifier (1.0–2.5). The resulting CRI provides an index of the cumulative biomechanical burden associated with a single posture window. Because TM is assigned from D, duration enters the index twice: once directly and once through the TM. For any given PS, CRI is therefore, by construction, a monotonically increasing function of duration. This property determines how the correlation results should be interpreted (Sections 4.4 and 5.3).

Risk Category (RC): A categorical interpretation of CRI values as Low, Medium, or High risk, derived through a dual-source approach combining pilot-anchored percentile thresholds and ISO 11226 angle-band criteria.

3.2. Theoretical Framework

The theoretical framework of MIERAT organises the tool's logic around three interconnected components, as summarised in Table 4.

Table 4. Logical components of the MIERAT theoretical framework: descriptions, theoretical bases, and expected contributions.

Component	Description	Theoretical Basis	Expected Contribution
Inputs	Observed body-segment postures, applied loads, and exposure durations collected in real-time automotive service operations.	Biomechanical Theory [16]; Exposure Theory [18]	Quantify posture-related strain and fatigue risk as a function of duration.
Processes	REBA-based posture scoring → time modifier application → CRI computation → Low/Medium/High risk categorisation.	Fatigue–Failure Theory [17]; Exposure Accumulation Models [30]	Transform raw observations into measurable, reproducible ergonomic indicators.
Outputs	Prioritised high-risk task lists; recommended ergonomic controls; PDCA-aligned feedback loops.	Systems Theory [20]; ISO 45001 Continuous Improvement [19]	Enable data-driven ergonomic decisions and structured workplace improvement.

Integration within MIERAT operates at three levels. At the theoretical level, biomechanical loading theory, exposure accumulation science, and systems-based occupational health and safety (OHS) management are unified into a coherent conceptual architecture. At the measurement level, integration is operationalised through the CRI formula, which combines REBA-derived posture severity, exposure duration, and fatigue-calibrated time modifiers into a single risk metric. At the operational level, CRI-based risk categories are mapped to ISO 45001 PDCA action protocols, so that assessment outputs can be translated into structured workplace improvements.

3.3. Conceptual Architecture: Seven Assessment Sections

MIERAT's conceptual architecture is organised into seven interrelated assessment sections: (1) administrative and task context profile; (2) posture episode identification through posture windowing; (3) REBA-based postural risk scoring; (4) exposure duration and time modifier assignment; (5) CRI computation; (6) body-region risk distribution analysis; and (7) risk interpretation and intervention planning. Together, these sections provide a structured pathway from contextual task documentation to actionable ergonomic decisions, aligned with the ISO 45001 PDCA cycle.

3.4. Process Flow

The MIERAT process flow provides a structured 12-step pathway from task scoping through posture scoring, duration capture, CRI computation, risk categorisation, internal consistency checking, and inter-rater reliability verification to the archiving of assessments. The flow incorporates explicit consistency and reliability checks, so that assessment outputs are systematically evaluated before they inform ergonomic decisions. A recalibration loop is triggered when inter-rater agreement falls below target thresholds (ICC < 0.90; $\kappa < 0.60$), supporting continuous improvement in observer consistency.

3.5. Design Principles

MIERAT's design was governed by six principles: (1) simplicity and accessibility—operable without specialised equipment in small and medium-sized enterprise (SME) settings; (2) accuracy and standardisation—posture scoring aligned with REBA and ISO 11226/11228; (3) reliability and repeatability—structured posture windows, observer training guide, and ICC/Kappa validation protocols; (4) integration with ISO safety management via PDCA; (5) practical relevance to automotive service work through posture

windowing of dynamic tasks; and (6) decision-support orientation through CRI threshold-triggered intervention recommendations.

3.6. Worked Example: MIERAT Scoring Process

To illustrate how these principles translate into practice, consider posture window UVS-04a (loosening the drain plug during under-vehicle servicing). The observer first identifies the posture episode boundary (from initial reach to plug release) and scores it using standard REBA body-region criteria, yielding a REBA score of 9 (high posture severity, driven by trunk rotation and an awkward reach angle). The posture is then timed with a stopwatch and checked against video timestamps, giving a duration of 191 s. Because this duration exceeds 120 s, the highest Time Modifier band applies (TM = 2.5; Table 2). The Cumulative Risk Index is then computed as CRI = PS × D × TM = 9 × 191 × 2.5 = 4297.5, placing this posture window in the High-risk category. By comparison, a REBA-only assessment of the same posture window would report only the static score of 9, with no indication that the posture was sustained for more than 3 min. This comparison illustrates what MIERAT adds to conventional static tools such as REBA.

4. Materials and Methods

4.1. Study Setting and Task Categories

Data collection was conducted at the Mabopane Automotive Centre (MAC), a small- to medium-scale independent vehicle service workshop in Mabopane, Pretoria, Gauteng Province, South Africa. The site was selected because it exemplifies the resource-constrained operational conditions for which MIERAT was designed: limited automation, heavy reliance on manual task execution, and the absence of formal ergonomic risk assessment protocols. Three core automotive service task categories were selected for observation: (1) under-vehicle servicing (oil change, fluid checks, under-vehicle inspection); (2) tyre fitment (wheel removal, tyre demounting, mounting, wheel balancing, and refitting); and (3) brake and suspension repair (brake pad replacement, caliper removal, suspension disassembly and reassembly, wheel alignment). These categories were chosen because together they represent the most ergonomically demanding operations in small workshops and provide sufficient variability in posture intensity, force demand, and exposure duration to test the tool across a range of conditions. Table 5 lists the constituent steps of each task category in process-flow order, to support reproducibility in workshops where the same tasks may be performed with minor procedural variations. Because workshop and site photographs were not covered by the ethics approval or the data collection protocol, Table 5 serves as a text-based alternative.

Table 5. Constituent steps of each task category, in process-flow order.

Task Category	Constituent Steps
Under-vehicle servicing	Wheel chocking bend → *transition* → guide vehicle alignment → operate lift controls → monitor lift (standing) → *transition* → under-vehicle entry posture → visual scan across components → *transition*→ loosen drain plug → passive drain flow hold → replug and wipe area → *transition*→ remove oil filter (overhead reach) → install new filter → component checks (squat reach) → tighten overhead fasteners → *transition* → refill fluids (lean over bay) → cap and dipstick check → leak check underneath → torque-check critical fasteners → *transition* → lower vehicle (control press) → remove stands/check bay clear → cleanup tools and area → documentation/job card

Table 5. *Cont.*

Task Category	Constituent Steps
Tyre fitment	Vehicle reception and apply wheel chocks → *transition* → pre-inspection of tyres → loosen wheel nuts (forceful) → *transition* → vehicle lifting → wheel removal → tyre deflation and valve removal → bead breaking → tyre demounting → rim preparation → tyre mounting → inflation and bead seating → wheel balancing and refit → final torque and documentation
Brake and suspension repair	Safety and setup → *transition* → symptom scan → *transition* → baseline visual inspection → brake diagnosis → *transition* → suspension diagnosis → repair plan and parts review → brake component removal → suspension disassembly → *transition* → component replacement → reassembly and torque → fluid service and bleeding → wheels and alignment (all four wheels)

Note: Italicised *transition* denotes a brief movement between consecutive working steps; MIERAT times and scores each transition as a separate posture window (e.g., T-01, TF-01b and BSR-01T in Section 5.2).

4.2. Sampling and Data Collection

A census sampling approach was adopted to include all relevant automotive service tasks within the defined study scope. The unit of analysis was the task episode, defined as a discrete work cycle from first tool contact to task completion. A total of 58 task episodes were observed across the three task categories. Three trained observers each independently scored the same set of task episodes, giving three independent scoring passes per task category and pooled sample sizes of 81 posture-window observations for under-vehicle servicing, 45 for tyre fitment, and 48 for brake and suspension repair. These pooled observations were used for the reliability and correlation analyses reported in Section 5. Data collection followed a fixed sequence for each task episode: task and posture-window identification, REBA-based posture severity scoring, stopwatch-based duration timing with video timestamp validation, Time Modifier band assignment, and CRI computation. The other two observers then cross-checked each episode independently to support the inter-rater reliability analysis. This sample size follows established guidance for observational reliability studies, which prioritise measurement depth and precision over sampling breadth [31,32]. Multiple posture windows were captured within each task episode, yielding a substantially larger effective dataset for CRI computation and reliability analysis than the episode count alone would suggest. Because the unit of analysis was the task episode rather than the individual worker, worker identity was not retained in a form that would allow unique contributors to be counted; the potential effect of repeated contributions by the same technicians on the statistical independence of observations is acknowledged as a limitation in Section 6.5. Data were collected over a four-week period, with sessions scheduled during regular working hours to preserve ecological validity. Only workers with $\geq$1 year of experience in the observed task area were included, and tasks performed during abnormal or non-routine operations were excluded so that the data reflected competent, representative work practices.

For each posture window, observers recorded the following: posture ID, task phase, body regions engaged, REBA-based posture severity score (PS), start and end timestamps, duration in seconds (D), time modifier category (TM), computed CRI value, and qualitative observational notes. Duration was measured with a digital stopwatch and validated against video timestamps to ensure accuracy.

4.3. Observer Training

Before data collection, all three observers completed calibration training based on the MIERAT Observer Training Guide, a structured protocol covering training objectives,

observation principles (e.g., focusing on posture transitions and avoiding estimated or guessed durations), posture-episode identification rules, the timing protocol, REBA scoring best practice, time modifier assignment rules, and CRI computation. Observers scored each task episode independently and did not have access to one another's REBA, duration, or CRI ratings during data collection; scores were compared only after the fieldwork was complete, for the inter-rater reliability analysis reported in Sections 5.4 and 5.5. Because all three observers completed the same calibration training, full blinding to the study's underlying duration-risk hypothesis was not feasible; this shared training exposure is acknowledged as a limitation in Section 6.5.

4.4. Statistical Analysis

Four distinct analyses were performed to evaluate the tool. Content validity was assessed through an expert panel review using the Content Validity Index (I-CVI and S-CVI/Ave) methodology [33–35]. The internal consistency of the CRI construct was examined using Spearman's rank correlation coefficient (ρ) for the associations between REBA scores, posture duration, and CRI values. Because posture severity and duration are inputs to the CRI equation, and duration additionally determines the TM, correlations between CRI and these variables are mathematically expected. They were therefore interpreted only as a check that the index behaves as designed, and not as independent evidence of construct validity. The REBA–Duration correlation, which involves two separately measured inputs, was examined to establish whether severity and duration convey distinct information. Spearman's ρ was selected because REBA scores are ordinal and CRI distributions may deviate from normality [36]. Correlation strength was interpreted as: weak (0.10–0.39), moderate (0.40–0.69), strong (0.70–0.89), and very strong (0.90–1.00) [37]. For each ρ, 95% confidence intervals were computed using the Fisher z-transformation, based on the pooled number of posture-window observations per task category across the three independent observer scoring passes (n = 81 for under-vehicle servicing, 45 for tyre fitment, 48 for brake and suspension repair). Inter-rater reliability of continuous CRI scores was evaluated using the Intraclass Correlation Coefficient (ICC (2,1)), a two-way random-effects model with absolute agreement and single-rater measurement. This model is appropriate when every observer rates the same subjects and the observers are regarded as representative of the wider population of trained observers who may apply the tool [31,38]. For absolute agreement, the two-way random-effects and two-way mixed-effects models use the same formula and give the same coefficient, so the choice between them affects how far the result can be generalised and not its value [31]. ICC values were interpreted as: poor (<0.50), moderate (0.50–0.75), good (0.75–0.90), and excellent (>0.90). Inter-rater agreement for categorical risk classifications (Low, Medium, High) was evaluated using Weighted Cohen's Kappa (κ) [39], interpreted according to Landis and Koch [40]: slight (0.00–0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80), and almost perfect (0.81–1.00). Data transcription and CRI computation were performed in Microsoft Excel (Microsoft 365 MSO Version 2608, Build 16.0.20326.20072; Microsoft Corporation, Redmond, WA, USA), and all inferential statistical analyses (Spearman's ρ, ICC, Weighted Cohen's κ) were conducted using IBM SPSS Statistics Version 29 (Build 29.0.2.0; IBM Corp., Armonk, NY, USA).

4.5. Ethical Considerations

Ethical approval was obtained from the University of Johannesburg Research Ethics Committee prior to data collection. Informed consent was obtained from all participants. Anonymity was maintained through participant coding and the blurring of identifying features in video recordings. Data were stored in encrypted databases accessible only to

the research team. Observations were non-intrusive and did not interfere with normal work activities.

5. Results

5.1. Content Validity

Content validity was assessed through review by a panel of four subject-matter experts (professional ergonomists, occupational therapists, and occupational health and safety officials familiar with automotive-sector work), whose combined expertise spanned ergonomics, industrial engineering, occupational health and safety, and human factors. A panel of four exceeds the minimum of three independent experts required to meet the CVI threshold of 0.78 and to distinguish genuine consensus from chance agreement reliably, according to the binomial probability analysis of Polit and Beck [34] and consistent with the CVI methodology of Lynn [33].

The panel rated ten MIERAT component items using a four-point relevance scale, with ratings of 3 ('quite relevant') or 4 ('highly relevant') treated as expert agreement. Table 6 presents the item-level CVI (I-CVI) and expert rating outcomes for each component.

Table 6. Content Validity Index (CVI) results for MIERAT components based on four expert ratings.

Item	MIERAT Component	Experts Rating 3 or 4 (*n*)	I-CVI	Decision
1	REBA posture scoring dimensions	4/4	1.00	Retained
2	Task description and posture identification fields	4/4	1.00	Retained
3	Body region observed	4/4	1.00	Retained
4	Start-time and end-time recording	4/4	1.00	Retained
5	Duration calculation	4/4	1.00	Retained
6	Time modifier scale	3/4	0.75	Revised
7	Cumulative Risk Index calculation logic	4/4	1.00	Retained
8	Low, Medium, and High risk classification	4/4	1.00	Retained
9	Observation notes and contextual field	3/4	0.75	Revised
10	Overall usability of the observation sheet	4/4	1.00	Retained
	Scale-level S-CVI/Ave		0.95	Strong content validity confirmed

Eight of the ten components achieved an I-CVI of 1.00, reflecting unanimous expert agreement. Two components, the time modifier scale (Item 6) and the observation notes field (Item 9), achieved I-CVI values of 0.75. Both were retained after revision: the explanatory notes for the time modifier scale were strengthened to clarify how the duration bands were calibrated, and the observation notes field was kept because it captures contextual information, such as confined-space constraints and tool-handling difficulty, that cannot be represented numerically. The overall S-CVI/Ave of 0.95 exceeds the recommended threshold of 0.90 [34], indicating strong content validity across the instrument.

5.2. Task-Level CRI Profiling: Descriptive Statistics

For all three task categories, CRI values were computed for each posture window and summarised as means, standard deviations, medians, and interquartile ranges across the three observations. Table 7 summarises the tasks with the highest mean CRI in each category, together with the corresponding reliability statistics.

Table 7. Tasks with the highest mean CRI in each task category, with descriptive statistics and reliability indices.

Task Category	Highest Mean CRI Task	Mean CRI (Highest)	SD	Median CRI	ICC (2,1)	Kappa Range (κ)
Under-vehicle servicing	UVS-04a (Drain plug loosening)	4297.5	0.00	4297.5	0.993	0.557–0.772
Under-vehicle servicing	UVS-03a (Under-vehicle entry)	3757.5	0.00	3757.5	0.993	0.557–0.772
Tyre fitment	TF-08 (Tyre demounting)	31,955.0	4565.00	31,955.0	0.975	0.557–0.722
Tyre fitment	TF-12 (Wheel balancing & refit)	30,950.0	2680.35	32,497.5	0.975	0.557–0.722
Brake & suspension	BSR-12 (Wheels & alignment)	15,326.7	50.3	15,320.0	0.997	0.893–1.000
Brake & suspension	BSR-08 (Suspension disassembly)	5952.5	415.3	5737.5	0.997	0.893–1.000

5.2.1. Under-Vehicle Servicing

The task with the highest CRI in under-vehicle servicing was UVS-04a (loosening the drain plug), which produced the highest values in all three observations (mean CRI = 4297.5; SD = 0.00; median = 4297.5). This task combines high posture severity (REBA = 9) with sustained forceful torque application for 191 s, yielding a time modifier of 2.5. The under-vehicle entry posture (UVS-03a) produced the second-highest mean CRI (3757.5), driven by pronounced neck extension and trunk flexion sustained for 167 s. Despite only medium posture severity (REBA = 5), the passive drain flow hold (UVS-04b) generated a mean CRI of 3612.5 because of its extended duration (289 s). Under the CRI model, low-exertion holding postures can therefore accumulate an exposure score comparable to that of forceful actions when they are prolonged. Component checks in squat reach (UVS-06a) produced a mean CRI of 3460.0 (SD = 0.00), identifying deep squatting as a consistent hotspot. Transition tasks and short control operations yielded negligible CRI values (e.g., T-01: mean CRI = 14.0), indicating that the tool separates sustained working postures from brief recovery movements.

5.2.2. Tyre Fitment

Tyre fitment produced substantially higher CRI values than under-vehicle servicing, reflecting the longer task durations and heavier manual handling demands of wheel removal, tyre manipulation, and wheel refitting. Tyre demounting (TF-08) produced the highest mean CRI (31,955.0; SD = 4565.0), driven by extended lever manipulation with wrist deviation and trunk flexion sustained for 1826 s. Wheel removal (TF-05) produced a mean CRI of 30,825.0 (SD = 3082.5), peaking at 33,907.5 in Observation 3, when deeper forward flexion was recorded. Tyre mounting (TF-10) and wheel balancing and refitting (TF-12) produced comparable means of 30,479.2 and 30,950.0, respectively, both exceeding 25,000 across all observations.

Loosening wheel nuts (TF-03a) generated a mean CRI of 24,960.0 (SD = 3120.0), driven by forceful torque application and trunk twisting sustained for 1248 s. Transitional tasks (TF-01b; TF-03b) produced CRIs of 158.4 and 126.0, respectively, indicating a negligible contribution to cumulative exposure.

5.2.3. Brake and Suspension Repair

Wheels and alignment (BSR-12) was the dominant ergonomic hotspot in brake and suspension repair, producing a mean CRI of 15,326.7 (SD = 50.3; median = 15,320.0). The very small standard deviation shows that this task was scored as high risk most consistently across observations. Its CRI was driven by prolonged lifting, squatting, and repeated

alignment adjustments across four wheels, sustained for approximately 765 s at a time modifier of 2.5. Suspension disassembly (BSR-08) was the second-highest contributor, with a mean CRI of 5952.5 (SD = 415.3), reflecting heavy strut lifting under combined trunk flexion and shoulder loading. Brake diagnosis (BSR-04; mean CRI = 3747.5), suspension diagnosis (BSR-05; mean CRI = 3652.5), reassembly and torque (BSR-10; mean CRI = 3589.2), and brake component removal (BSR-07; mean CRI = 3494.2) together formed a high-risk cluster, each driven by combinations of kneeling, trunk flexion, and forceful torque application. Transition tasks (BSR-01T, BSR-02T, BSR-04T, BSR-08T) produced mean CRIs below 102, again indicating a negligible contribution.

5.3. Internal Consistency of the CRI Construct: Spearman's Rank Correlations

Spearman's rank correlation was used to check whether CRI values behaved as designed, that is, whether they increased monotonically with posture severity and exposure duration. Table 8 presents the correlation coefficients, confidence intervals, and significance levels for all three task categories.

Table 8. Spearman's rank correlation coefficients (with 95% confidence intervals) between REBA score, posture duration, and CRI across all three task categories. Correlations were computed on pooled posture-window observations across the three independent observer scoring passes per category (n = 81 for under-vehicle servicing, 45 for tyre fitment, 48 for brake and suspension repair); $p < 0.001$ for all.

Task Category	Variable Pair	Spearman ρ	*p*-Value	95% CI	Interpretation
Under-vehicle servicing	REBA vs. Duration	0.596	<0.001	[0.434, 0.721]	Moderate
Under-vehicle servicing	REBA vs. CRI	0.856	<0.001	[0.784, 0.905]	Strong
Under-vehicle servicing	Duration vs. CRI	0.910	<0.001	[0.863, 0.941]	Very strong
Tyre fitment	REBA vs. Duration	0.875	<0.001	[0.782, 0.930]	Strong
Tyre fitment	REBA vs. CRI	0.939	<0.001	[0.891, 0.966]	Very strong
Tyre fitment	Duration vs. CRI	0.961	<0.001	[0.930, 0.979]	Very strong
Brake & suspension	REBA vs. Duration	0.827	<0.001	[0.710, 0.900]	Strong
Brake & suspension	REBA vs. CRI	0.940	<0.001	[0.895, 0.966]	Very strong
Brake & suspension	Duration vs. CRI	0.945	<0.001	[0.903, 0.969]	Very strong

Note: Interpretation bands: weak (0.10–0.39), moderate (0.40–0.69), strong (0.70–0.89), very strong (0.90–1.00).

Across all three task categories, Duration–CRI correlations were the strongest of the three variable pairs, ranging from ρ = 0.910 (under-vehicle servicing) to ρ = 0.961 (tyre fitment). These correlations must be interpreted with caution. Duration enters the CRI equation directly (CRI = PS × D × TM) and also determines the Time Modifier, so CRI is, by construction, a monotonically increasing function of duration for any given posture severity score. A strong Duration–CRI correlation is therefore mathematically expected and does not constitute independent evidence of construct validity. It shows only that the index behaves as designed and reflects the weight given to duration in the formula; it should not be read as evidence that duration is the predominant real-world determinant of ergonomic burden. The same caveat applies to the REBA–CRI correlations (ρ = 0.856–0.940), because posture severity is likewise an input to the index. The REBA–Duration correlation is the only one of the three that involves two separately measured inputs. For under-vehicle servicing, it was moderate (ρ = 0.596), indicating that severity and duration are not collinear in this task category and that each conveys distinct information. It was stronger for tyre fitment (ρ = 0.875) and brake and suspension repair (ρ = 0.827), reflecting the more homogeneous posture-duration structure of these longer-cycle tasks. All coefficients were statistically significant ($p < 0.001$) and ranged from moderate to very strong (ρ = 0.596–0.961; 95% confidence intervals in Table 8). Taken together, these results support

the internal consistency of the CRI, but they cannot establish its construct or criterion validity. That requires testing CRI against outcomes measured independently of the index, such as musculoskeletal symptoms, fatigue, biomechanical loading, or injury data (Sections 6.1 and 6.5).

5.4. Inter-Rater Reliability: Intraclass Correlation Coefficients

The ICC (2,1) model was applied using ANOVA-derived variance components to assess inter-rater reliability for continuous CRI scores across the three observers. Table 9 presents the ICC values, confidence intervals, reliability classifications, and ANOVA between-subjects significance for each task category.

Table 9. Intraclass Correlation Coefficient (ICC (2,1)) results for MIERAT CRI scores across three task categories, with 95% confidence intervals and ANOVA significance.

Task Category	*n* (Task Episodes)	ICC (2,1)	95% CI	Reliability Level	ANOVA: Between-Subjects F (*p*)
Under-vehicle servicing	27	0.993	[0.989, 0.996]	Excellent	$F = 448.36$, $p < 0.001$
Tyre fitment	15	0.975	[0.960, 0.994]	Excellent	$F = 173.75$, $p < 0.001$
Brake & suspension	16	0.997	[0.994, 0.999]	Excellent (near perfect)	$F = 294.29$, $p < 0.001$

All three ICC values exceeded 0.97, indicating excellent reliability [31,41]. The ICC of 0.993 for under-vehicle servicing indicates that virtually all variance in CRI scores was attributable to differences between task episodes rather than to observer inconsistency. ANOVA between-subjects F-values were highly significant across all categories ($p < 0.001$), indicating that CRI scores differed systematically between task episodes. This shows that the tool separates episodes on the CRI scale; it does not in itself show that these differences correspond to differences in musculoskeletal outcomes. Between-rater effects were non-significant for under-vehicle servicing ($F = 2.21$, $p = 0.120$) and brake and suspension repair ($F = 0.017$, $p = 0.983$), indicating that observers applied the scoring protocol with negligible systematic bias.

A significant between-rater effect was detected for tyre fitment ($F = 7.97$, $p = 0.002$), suggesting some observer variability in this more dynamic task category; however, given the excellent ICC of 0.975, this effect did not meaningfully compromise reliability. These results indicate that trained observers applied the structured combination of REBA scoring, posture-window timing, and TM assignment consistently.

5.5. Inter-Rater Agreement: Cohen's Kappa

Weighted Cohen's Kappa was computed for all pairwise observer comparisons within each task category to assess categorical risk classification agreement. Table 10 presents the Kappa values and agreement classifications.

Brake and suspension tasks produced the strongest categorical agreement, with $\kappa = 1.000$ between Observers 1 and 3 (perfect agreement) and $\kappa = 0.893$ for the remaining pairings (near-perfect agreement). Under-vehicle servicing and tyre fitment produced moderate-to-substantial agreement ($\kappa = 0.557$–0.772), with the lowest κ values consistently appearing in the Observer 2 vs. Observer 3 pairing. This pattern probably reflects the influence of borderline CRI values near category thresholds, where minor scoring differences produce categorical disagreements. In tyre fitment, the dynamic and force-variable nature of lever-based tasks (e.g., tyre demounting, bead breaking) increases the probability of such borderline disagreements. All κ values exceeded 0.50 and therefore fell within or above the 'moderate' band of Landis and Koch [40]. However, the two Observer 2 vs. Observer 3 pairings with $\kappa = 0.557$ fell below MIERAT's own target of $\kappa \geq 0.60$ and would trigger the

recalibration loop described in Section 3.4. Because scores were compared only after the fieldwork was complete (Section 4.3), this recalibration was not applied during data collection. The strongest agreement was consistently observed for tasks with clear, unambiguous risk profiles (e.g., BSR-12 wheels and alignment; UVS-04a drain plug loosening), where CRI values are far from category boundaries.

Table 10. Weighted Cohen's Kappa (κ) results for categorical CRI risk classification (Low/Medium/High) across three task categories and three observer pairings.

Task Category	Observer Pairing	Cohen's κ	Agreement Level	Landis & Koch Classification
Under-vehicle servicing	Obs 1 vs. Obs 2	0.772	Substantial	Substantial Agreement
Under-vehicle servicing	Obs 1 vs. Obs 3	0.722	Substantial	Substantial Agreement
Under-vehicle servicing	Obs 2 vs. Obs 3	0.557	Moderate	Moderate Agreement
Tyre fitment	Obs 1 vs. Obs 2	0.722	Substantial	Substantial Agreement
Tyre fitment	Obs 1 vs. Obs 3	0.682	Substantial	Substantial Agreement
Tyre fitment	Obs 2 vs. Obs 3	0.557	Moderate	Moderate Agreement
Brake & suspension	Obs 1 vs. Obs 2	0.893	Near-perfect	Almost Perfect Agreement
Brake & suspension	Obs 1 vs. Obs 3	1.000	Perfect	Perfect Agreement
Brake & suspension	Obs 2 vs. Obs 3	0.893	Near-perfect	Almost Perfect Agreement

6. Discussion

6.1. Validation Summary and Tool Performance

Taken together, the results provide initial support for MIERAT as a content-valid, reliable, and practically deployable ergonomic assessment instrument; they do not yet establish its construct or criterion validity. The S-CVI/Ave of 0.95 indicates that expert raters with diverse but complementary backgrounds in ergonomics, occupational health, and industrial engineering regard the tool's components as relevant and comprehensive. The Spearman correlations provide evidence of internal consistency for the CRI construct: in all three task categories, Duration–CRI correlations exceeded $\rho = 0.90$ ($p < 0.001$; see Table 8 for 95% CIs), showing that CRI behaves monotonically with its own inputs, as designed.

This is a necessary but not a sufficient condition for construct validity. Duration is included directly in the CRI equation and also determines the Time Modifier, so a strong Duration–CRI correlation is mathematically expected; correlating CRI with its own inputs cannot, by itself, constitute independent construct validation. The consistently stronger Duration–CRI relationship relative to REBA–CRI reflects the larger mathematical contribution of duration to the index as constructed, and nothing more should be inferred from it. Whether CRI, and the weight it assigns to exposure duration, corresponds to real-world ergonomic burden can be established only by testing CRI and its components against independent outcomes that are not derived from the index. Such outcomes include musculoskeletal symptoms and discomfort, perceived or physiological fatigue, directly measured biomechanical loading (e.g., electromyography or wearable inertial sensors), and injury records. Until such evidence is available, CRI should be interpreted as a structured index of cumulative postural exposure rather than as a validated predictor of musculoskeletal risk. This validation is the main priority for future work (Section 6.5).

The excellent ICC values (0.975–0.997) across all task categories show that MIERAT produces highly reproducible CRI scores when applied by trained observers. Reproducibility of this kind is essential for any tool intended for repeated use in audits, monitoring, and the evaluation of interventions. The near-perfect ICC for brake and suspension repair (0.997) is particularly noteworthy and suggests that the structured scoring protocol and

clear posture-window definitions standardise observer behaviour even in the most complex of the three task categories [31]. The weighted Cohen's Kappa values for categorical agreement ranged from moderate to perfect (0.557–1.000) [40], indicating that observers generally assigned CRI scores to the same risk categories; agreement was strongest for brake and suspension tasks. The lower Kappa values observed for certain tyre fitment and under-vehicle servicing pairings ($\kappa = 0.557$, below the tool's recalibration target of $\kappa \geq 0.60$; Section 3.4) point to the need for threshold calibration exercises in future training programmes, particularly for tasks where CRI values cluster near category boundaries.

6.2. Ergonomic Risk Profile Findings

The task-level CRI profiling indicates that CRI-estimated exposure in automotive service work is not distributed uniformly across activities; rather, it is concentrated in a small number of high-load, extended-duration tasks that account for most of the cumulative exposure. In all three task categories, transitions and machine-assisted operations yielded negligible CRI values, which shows that MIERAT can separate candidate ergonomic hotspots from low-scoring activities within the same workflow.

For under-vehicle servicing, the drain plug loosening sequence (UVS-04a and UVS-04b) is the highest-priority intervention target: its combined CRI exceeded 7900 over a 480-s window, driven by forceful torque application and prolonged passive holding in constrained under-vehicle postures. Notably, the passive drain flow hold (UVS-04b), with only medium posture severity (REBA = 5), generated a mean CRI of 3612.5, comparable to that of tasks with REBA scores of 8 or 9.

This pattern, which conventional snapshot tools would not capture, illustrates MIERAT's core theoretical proposition, consistent with cumulative fatigue–failure theory [26]: sustained low-intensity postures may accumulate an ergonomic burden comparable to that of short-duration, high-intensity postures, in which case risk management programmes that focus exclusively on posture severity would misidentify the primary drivers of cumulative strain. Because this pattern follows directly from the structure of the CRI formula, it should be treated as a model-based prediction that remains to be confirmed against independent outcome data.

For tyre fitment, the CRI values associated with wheel removal, tyre demounting, mounting, and balancing (mean CRIs of approximately 30,000–32,000) substantially exceed those recorded for under-vehicle servicing and brake and suspension tasks. This reflects the exceptionally long uninterrupted durations of these activities (1233–1857 s) combined with moderate-to-high posture severity, and identifies tyre fitment as the operation with the highest CRI-estimated cumulative exposure in the studied workshop.

These findings are consistent with ISO 11228-1 [22], which identifies the repetitive manual handling of objects heavier than 3 kg as a trigger for risk assessment; the mean tyre load in the study setting substantially exceeded this threshold.

For brake and suspension repair, the near-identical CRI values for wheels and alignment (BSR-12) across all three observations (15,280–15,380) identify this task as the most consistently scored hotspot in the brake and suspension workflow, driven by its very long duration at the highest TM band combined with sustained multi-joint loading. The concentration of CRI in a small cluster of high-load tasks (BSR-12, BSR-08, BSR-04, BSR-10, and BSR-07) provides clear guidance for prioritising ergonomic interventions.

6.3. Comparison with Existing Assessment Tools

MIERAT occupies a distinctive niche relative to existing ergonomic assessment tools. It extends REBA's validated whole-body posture scoring framework [9] by adding an explicit posture-duration dimension, which none of the five established tools reviewed

in Table 1 provides in this form, while retaining REBA's field practicability and modest training requirements. The present findings illustrate what a REBA score alone does not show: under the CRI model, a medium-severity posture sustained for an extended period (e.g., UVS-04b: REBA = 5, D = 289 s, CRI = 3612.5) may carry a cumulative exposure score close to that of a high-severity posture held for a shorter period (e.g., UVS-04a: REBA = 9, D = 191 s, CRI = 4297.5), whereas the REBA scores alone would rank the second posture far above the first. This posture-duration interaction, grounded in cumulative fatigue–failure theory, is the core theoretical contribution that distinguishes MIERAT from its predecessors. OCRA provides the most sophisticated temporal modelling among existing tools through its frequency and recovery time parameters, but its computational complexity and primary focus on upper-limb repetitive tasks make it impractical for the whole-body, dynamic character of automotive service operations and poorly suited to SME deployment [13]. MIERAT achieves temporal integration within a much simpler computational structure that is accessible to users without specialist expertise in ergonomics. This raises a natural question: if MIERAT explicitly incorporates exposure duration while REBA does not, do MIERAT's outputs differ systematically from, and potentially outperform, REBA-alone assessments? The theoretical expectation is that MIERAT should more effectively flag risk in tasks that combine moderate-to-high posture severity with prolonged duration, where a static REBA score would understate the cumulative burden, as illustrated by UVS-04b above. The present study provides indirect, correlational support for this expectation: the REBA–Duration correlation for under-vehicle servicing was only moderate ($\rho = 0.596$; Table 8), indicating that REBA scores alone do not fully capture the duration-related risk that CRI incorporates.

However, this manuscript does not include a direct statistical comparison of MIERAT and REBA-alone classifications on the same task episodes (e.g., a paired-samples test with an effect-size threshold). Such a comparison would show whether the two tools classify tasks differently; whether MIERAT's classifications are more accurate can be determined only against independent outcome data. This comparison was outlined in the original study design but has not yet been completed, and we identify it here explicitly as a priority for the next phase of this research programme.

6.4. Practical and Regulatory Implications

MIERAT's design directly addresses the implementation barriers that have historically limited ergonomic risk assessment in small automotive workshops in South Africa and in comparable low-resource industrial settings. The tool requires no specialised equipment, no digital infrastructure, and no advanced ergonomic training beyond the structured observer calibration programme described in this paper. Its outputs, the CRI and the associated risk categories, are expressed in terms that are accessible to workshop supervisors, safety officers, and occupational health practitioners without ergonomics expertise, and can therefore support decision-making.

The tool's alignment with the Occupational Health and Safety Act 85 of 1993, the South African Ergonomics Regulations of 2019, and ISO 45001 provides a structured documentation framework that supports legislative compliance. CRI outputs can inform ergonomic hazard registers, risk audit trails, and engagement with occupational health practitioners and labour inspectors, provided that they are interpreted as exposure-based screening indicators pending validation against health outcomes. Interventions suggested by the present findings include mechanical tyre lifts and wheel handling aids for tyre fitment; adjustable vehicle lift heights and creeper platforms for under-vehicle servicing; torque-assist tools and height-adjustable component supports for brake and suspension repair; and structured

task rotation and micro-break schedules across all three task categories to reduce sustained posture duration.

6.5. Limitations

Several limitations should be acknowledged. First, MIERAT does not directly measure applied force; posture severity serves as a biomechanical proxy that may underestimate risk in tasks involving high force application within relatively neutral postures. This limitation is illustrated directly by the wheel-nut loosening task in tyre fitment (TF-03a; REBA = 8, CRI = 24,960), where forceful torque application combined with trunk twisting drove cumulative strain that posture scoring alone captures only indirectly, through its secondary effect on trunk angle rather than through force itself. A supplementary force or exertion proxy (e.g., an ordinal perceived-exertion rating) is recommended as a future extension. Second, environmental stressors, including heat, vibration, and noise, fall outside the tool's scoring scope and may contribute to an occupational health burden that CRI does not capture. Third, the present validation was conducted at a single small workshop in South Africa; although this enhances ecological validity, it limits the immediate generalisability of CRI thresholds to other regional contexts or workshop configurations, and multi-site validation across diverse automotive workshops is identified as a priority for future work. Fourth, because the unit of analysis was the task episode rather than the individual worker, worker identity was not retained in the dataset, and the extent to which task episodes reflect repeated observations from the same technicians, with implications for the statistical independence of observations, could not be determined post hoc; a worker-level random effect is recommended for future, larger-sample validation.

Fifth, the three observers scored task episodes independently, but full blinding to the study's underlying duration-risk hypothesis was not feasible, since all observers completed the same MIERAT calibration training; this shared expectation may have contributed to the moderate (rather than substantial) Kappa values observed for the Observer 2 vs. Observer 3 pairing in under-vehicle servicing and tyre fitment ($\kappa = 0.557$). Sixth, the significant between-rater effect detected for tyre fitment ($F = 7.97$, $p = 0.002$), although not large enough to compromise overall reliability, indicates that additional observer calibration for highly dynamic, force-variable tasks would strengthen scoring consistency, and may itself partly reflect the force-scoring ambiguity noted above. Seventh, 95% confidence intervals could not be computed for the Weighted Cohen's Kappa values (Table 10), as the specific numeric thresholds used to classify CRI scores into Low, Medium, and High risk categories during the original analysis were not retained; only the resulting κ point estimates and their Landis and Koch interpretive bands remain available. Future work should archive threshold definitions alongside raw data to permit full retrospective reanalysis. Finally, and most importantly, the construct and criterion validity of CRI have not yet been tested against any outcome measured independently of the index. The correlations reported in Section 5.3 involve the index's own inputs and are mathematically expected, so they cannot serve this purpose. Future studies should therefore validate CRI against independent outcomes, including self-reported musculoskeletal symptoms and discomfort, perceived or physiological fatigue, directly measured biomechanical loading, and injury records, ideally in prospective, multi-site designs. Until such evidence is available, the CRI risk categories should be regarded as provisional screening bands rather than as validated predictors of musculoskeletal outcomes.

7. Conclusions

This paper presented the development and initial field validation of MIERAT, the Manual Integrated Ergonomics Risk Assessment Tool, designed for small-scale automo-

tive service environments. MIERAT extends conventional static posture assessment by integrating posture severity and temporal exposure into a single Cumulative Risk Index (CRI = PS × D × TM), grounded in biomechanical loading theory, cumulative fatigue–failure models, and internationally recognised ergonomic standards.

The results provide initial evidence of the tool's content validity and inter-rater reliability. An S-CVI/Ave of 0.95 indicated strong content validity. Spearman's correlations between duration and CRI (ρ = 0.910–0.961) showed that CRI behaves consistently with its design, in which temporal exposure and posture severity jointly determine the index. Because duration is an input to the CRI equation and also determines the Time Modifier, these correlations are mathematically expected and are not independent evidence of construct validity. ICC values of 0.975–0.997 across the three task categories indicated excellent inter-rater reliability, with non-significant between-rater effects for under-vehicle servicing and brake and suspension repair. Categorical agreement ranged from moderate to perfect (κ = 0.557–1.000), which broadly supports the tool's reliability for practical risk classification, although two observer pairings fell below the tool's recalibration target of $\kappa \geq 0.60$.

Task-level CRI profiling identified drain plug loosening, under-vehicle entry posture, and passive drain flow holding as the primary ergonomic hotspots in under-vehicle servicing; wheel removal, tyre demounting, mounting, and balancing as the dominant hotspots in tyre fitment; and wheel alignment as the highest-scoring episode in brake and suspension repair. A key finding is that medium-severity postures sustained over extended durations can generate CRI values comparable to, or exceeding, those of short-duration high-severity postures.

Conventional snapshot tools cannot detect this pattern, which has direct implications for intervention prioritisation if it is confirmed against independent outcomes.

MIERAT makes three main contributions: (1) a theoretical contribution, by operationalising posture-duration integration as a practical, field-applicable cumulative risk metric; (2) a methodological contribution, through the posture-window approach to capturing temporal exposure in dynamic industrial environments; and (3) a practical contribution, by delivering a cost-effective, ISO-aligned ergonomic assessment framework suited to resource-constrained South African automotive workshops. The most important next step is to validate CRI against independent outcomes, namely musculoskeletal symptoms, fatigue, biomechanical loading, and injury data. Future research should also undertake longitudinal validation of the tool across multiple sites and workshop configurations, calibrate CRI thresholds empirically against physiological and self-reported discomfort outcomes, and evaluate the tool's sensitivity to change following ergonomic interventions.

Author Contributions: Conceptualization, K.M.; methodology, K.M.; formal analysis, K.M.; writing—original draft preparation, K.M.; writing—review and editing, P.M.M.; analysis and review, P.M.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the University of Johannesburg Research Ethics Committee.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data are available upon reasonable request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. World Health Organization. Musculoskeletal Conditions. WHO Fact Sheet. 2022. Available online: https://www.who.int/news-room/fact-sheets/detail/musculoskeletal-conditions (accessed on 11 January 2026).
2. International Labour Organization. *The Prevention of Occupational Diseases*; ILO: Geneva, Switzerland, 2013.
3. Tamene, A.; Mulugeta, H.; Ashenafi, T.; Thygerson, S.M. Musculoskeletal disorders and associated factors among vehicle repair workers in Hawassa City, Southern Ethiopia. *J. Environ. Public Health* **2020**, *2020*, 9472357. [CrossRef] [PubMed]
4. Hailu, T.; Desye, B.; Engdaw, D. Prevalence and associated factors of work-related musculoskeletal disorders among workers in manufacturing industries: A systematic review and meta-analysis. *Arch. Environ. Occup. Health* **2023**, *78*, 2042.
5. Zhang, P.; Arora, N.; Xiong, J.; Dong, R.G.; Welcome, D.E. Analysis of work-related musculoskeletal disorders in manufacturing: A systematic review. *Int. J. Environ. Res. Public Health* **2023**, *20*, 2225.
6. Republic of South Africa. *Occupational Health and Safety Act, No. 85 of 1993*; Government Gazette: Pretoria, South Africa, 1993.
7. Mashwama, X.N.; Aigbavboa, C.; Thwala, D. An assessment of the challenges and barriers in the implementation of health and safety in the South African construction industry. *Procedia Eng.* **2019**, *196*, 735–742.
8. Republic of South Africa. *Ergonomics Regulations Under the Occupational Health and Safety Act*; Government Gazette: Pretoria, South Africa, 2019.
9. Hignett, S.; McAtamney, L. Rapid Entire Body Assessment (REBA). *Appl. Ergon.* **2000**, *31*, 201–205. [CrossRef] [PubMed]
10. McAtamney, L.; Corlett, E.N. RULA: A survey method for the investigation of work-related upper limb disorders. *Appl. Ergon.* **1993**, *24*, 91–99. [CrossRef] [PubMed]
11. Karhu, O.; Kansi, P.; Kuorinka, I. Correcting working postures in industry: A practical method for analysis. *Appl. Ergon.* **1977**, *8*, 199–201. [CrossRef] [PubMed]
12. Li, G.; Buckle, P. Current techniques for assessing physical exposure to work-related musculoskeletal risks, with emphasis on posture-based methods. *Ergonomics* **1999**, *42*, 674–695. [CrossRef] [PubMed]
13. Occhipinti, E. OCRA: A concise index for the assessment of exposure to repetitive movements of the upper limbs. *Ergonomics* **1998**, *41*, 1290–1311. [CrossRef] [PubMed]
14. Ma, L.; Chablat, D.; Bennis, F.; Zhang, W.; Guillaume, F. A new ergonomic assessment method for real-time monitoring of workers in dynamic work environments. *Ergonomics* **2022**, *65*, 423–437.
15. Punnett, L.; Wegman, D.H. Work-related musculoskeletal disorders: The epidemiologic evidence and the debate. *J. Electromyogr. Kinesiol.* **2004**, *14*, 13–23. [CrossRef] [PubMed]
16. Marras, W.S.; Karwowski, W. *Fundamentals and Assessment Tools for Occupational Ergonomics*, 2nd ed.; CRC Press: Boca Raton, FL, USA, 2006.
17. Gallagher, S.; Heberger, J.R. Examining the interaction of force and repetition on musculoskeletal disorder risk: A systematic literature review. *Hum. Factors* **2013**, *55*, 108–124. [PubMed]
18. Mathiassen, S.E. Diversity and variation in biomechanical exposure: What is it, why does it matter and how is it quantified? *Ergonomics* **2006**, *49*, 1011–1021.
19. *ISO 45001*; Occupational Health and Safety Management Systems—Requirements with Guidance for Use. International Organization for Standardization: Geneva, Switzerland, 2018.
20. Checkland, P. *Systems Thinking, Systems Practice*; Wiley: Chichester, UK, 1999.
21. *ISO 11226*; Ergonomics—Evaluation of Static Working Postures. International Organization for Standardization: Geneva, Switzerland, 2000.
22. *ISO 11228-1*; Ergonomics—Manual Handling—Part 1: Lifting, Lowering and Carrying. International Organization for Standardization: Geneva, Switzerland, 2021.
23. *ISO/TS 20646*; Ergonomics Guidelines for the Optimization of Musculoskeletal Workload. International Organization for Standardization: Geneva, Switzerland, 2014.
24. *ISO 11228-3*; Ergonomics—Manual Handling—Part 3: Handling of Low Loads at High Frequency. International Organization for Standardization: Geneva, Switzerland, 2007.
25. *ISO/TR 12295*; Ergonomics—Application Document for ISO Standards on Manual Handling (ISO 11228-1, ISO 11228-2 and ISO 11228-3) and Evaluation of Static Working Postures (ISO 11226). International Organization for Standardization: Geneva, Switzerland, 2014.
26. Gallagher, S.; Schall, M.C. Musculoskeletal disorders as a fatigue failure process: Evidence, implications and research needs. *Ergonomics* **2017**, *60*, 255–269. [CrossRef] [PubMed]
27. Hedge, A. *Ergonomic Workplace Design for Health, Wellness, and Productivity*; CRC Press: Boca Raton, FL, USA, 2016.
28. Schall, M.C.; Sesek, R.F.; Cavuoto, L.A. Fatigue in the workplace: A comparison of fatigue assessment methods in manufacturing. *Occup. Ergon.* **2020**, *15*, 91–106.
29. Mdlalose, K. Development of an Integrated Ergonomics Risk Assessment Tool for Automotive Service Workers. Doctoral Thesis, University of Johannesburg, Johannesburg, South Africa, 2026.

30. Kumar, S. Theories of musculoskeletal injury causation. *Ergonomics* **2001**, *44*, 17–47. [CrossRef]
31. Koo, T.K.; Li, M.Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **2016**, *15*, 155–163. Erratum in *J. Chiropr. Med.* **2017**, *16*, 346. [CrossRef] [PubMed]
32. Portney, L.G.; Watkins, M.P. *Foundations of Clinical Research: Applications to Evidence-Based Practice*, 4th ed.; F.A. Davis: Philadelphia, PA, USA, 2025.
33. Lynn, M.R. Determination and quantification of content validity. *Nurs. Res.* **1986**, *35*, 382–385. [CrossRef]
34. Polit, D.F.; Beck, C.T. The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Res. Nurs. Health* **2006**, *29*, 489–497. [CrossRef] [PubMed]
35. Yusoff, M.S.B. ABC of content validation and content validity index calculation. *Educ. Med. J.* **2019**, *11*, 49–54. [CrossRef]
36. Field, A. *Discovering Statistics Using IBM SPSS Statistics*, 5th ed.; SAGE Publications: London, UK, 2018.
37. Schober, P.; Boer, C.; Schwarte, L.A. Correlation coefficients: Appropriate use and interpretation. *Anesth. Analg.* **2018**, *126*, 1763–1768. [CrossRef] [PubMed]
38. Bartlett, J.W.; Frost, C. Reliability, repeatability and reproducibility: Analysis of measurement errors in continuous variables. *Ultrasound Obstet. Gynecol.* **2008**, *31*, 466–475. [CrossRef] [PubMed]
39. McHugh, M.L. Interrater reliability: The kappa statistic. *Biochem. Medica* **2012**, *22*, 276–282. [CrossRef]
40. Landis, J.R.; Koch, G.G. The measurement of observer agreement for categorical data. *Biometrics* **1977**, *33*, 159–174. [CrossRef]
41. Cicchetti, D.V. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol. Assess.* **1994**, *6*, 284–290. [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.