International Journal of
Technology, Public Administration, and Education
Vol. 2 No. 1. June 2026, pp. 17-22

A Cross-Instrument Reliability Analysis of Tems Pocket and Nemo Analyze in 4G LTE Walk Testing

Sufianti Munirman[1*], Apriana Toding[2], Riesa Krisna Astuti Sakir[3]
[1,3]Department of Electrical Engineering, Ujung Pandang State Polytechnic, South Sulawesi, Indonesia
[2]Department of Electrical Engineering, Universitas Kristen Indonesia Paulus, South Sulawesi, Indonesia

ORCIDs/Email:
First AUTHOR : https://orcid.org/0009-0003-6604-1964 / sufiantimunirman@poliupg.ac.id
Second AUTHOR: https://orcid.org/000-0002-8668-1729 / apriana.toding@ukipaulus.ac.id
Third AUTHOR : https://orcid.org/0009-0009-6143-9598 / riesakrisna@poliupg.ac.id

Corresponding author email: sufiantimunirman@poliupg.ac.id

Article Info
Recieved: 18/01/2026
Revised: 07/04/2026
Accepted: 25/05/2026
Online: 30/06/2026

Abstract
Walk test conclusions are only as trustworthy as the instruments that produce them, yet studies rarely test whether two measurement tools yield the same verdict on the same network. This study evaluates the cross-instrument reliability of Tems Pocket, analysed in TEMS Investigation, and Nemo Handy, analysed in Nemo Analyze, by measuring the 4G LTE networks of Telkomsel and XL Axiata along an identical route at Cimory Dairyland, a hillside tourist area in Gowa Regency, South Sulawesi, and comparing their classifications of RSRP, RSRQ, SINR, RSSI, and throughput. The two chains agreed strongly on the spatial distribution of every parameter, both locating the weakest signal in the same vegetated southwestern sector, but diverged on absolute category percentages, most sharply for RSRP, where one chain classified 80.34 per cent of a route as poor while the other classified the majority as normal. The novelty of this study is a reliability rule derived from this pattern: walk test conclusions are trustworthy at the level of spatial pattern but not at the level of absolute percentage from a single instrument, so multi-tool corroboration and the reporting of spatial rather than absolute findings should be standard practice, particularly when instrument condition, such as the ageing of the Tems Pocket unit used here, cannot be assumed equal.

Keywords: Walk Test, Measurement Reliability, Tems Pocket, Nemo Analyze, Instrument Validity

© 2026 by the author(s)
This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

INTRODUCTION

Every empirical conclusion about network performance rests on the instrument that produced the measurement, yet the instrument is often treated as a transparent window rather than as a potential source of variation. In 4G Long Term Evolution (LTE) walk testing, radio parameters, namely Reference Signal Received Power (RSRP), Reference Signal Received Quality (RSRQ), Signal to Interference plus Noise Ratio (SINR), Received Signal Strength Indicator (RSSI), and throughput, are logged by specialised hardware and software along the route (Krasniqi et al., 2015; Kurniawan &

Prasetyo, 2020; Herina, 2022). Different tools, however, may apply different calibrations, sampling behaviours, and classification boundaries, so that the same network measured by two tools need not yield the same numerical verdict.

The Indonesian walk test literature has used a range of tools, from G-NetTrack Pro (Efriyendro & Rahayu, 2017; Sukiswo & Zahra, 2018) to Genex Assistant (Karo et al., 2019) and Tems Pocket alongside comparison software (Dewa et al., 2016), and studies of highland and vegetated terrain have documented the propagation conditions such sites impose (Rahmania, 2020). What this body of work has largely left unexamined is the agreement between tools: when a study reports that a route is, say, sixty per cent in a given quality class, the reader cannot know whether a second tool would have returned the same figure. Without such a check, absolute percentages are reported with a confidence the underlying measurement may not warrant.

This study makes that check its explicit subject. By measuring the same networks along the same route with two independent chains, Tems Pocket analysed in TEMS Investigation and Nemo Handy analysed in Nemo Analyze, it evaluates where the two tools agree and where they diverge, across all five parameters. The novelty pursued is a reliability rule for walk test practice: a principled statement of which level of a walk test result, spatial pattern or absolute percentage, can be trusted from a single instrument, and what practice should follow when instruments cannot be assumed equivalent. The question is captured in the title: can two tools be trusted equally, and if not, at what level of the result does the divergence matter.

Sources of Divergence Between Measurement Tools

Two walk test tools measuring the same radio environment can diverge for several reasons that are independent of the network itself. Calibration differences shift the absolute values a tool reports for a given physical signal; different sampling rates and averaging windows change how transient conditions are captured; and different category boundaries, or different mappings of raw values to quality classes, move samples across class thresholds even when the underlying measurement is identical. Hardware condition adds a further source: an ageing handset may report systematically shifted values as its radio front end drifts from factory calibration (Herina, 2022; Dewa et al., 2016). Any of these can produce different absolute percentages from the same network.

Study Design for a Cross-Instrument Comparison

Crucially, these sources of divergence act mainly on absolute values, not on spatial structure. The relative ordering of locations, which points along a route are stronger or weaker, is driven by the physical propagation environment, which both tools observe in common; a location shadowed by terrain and screened by vegetation will read as comparatively weak on any correctly functioning tool, even if the two tools disagree on the exact decibel value or class label (Rahmania, 2020; Karo et al., 2019). This distinction, between absolute agreement and spatial agreement, is the analytical hinge of the study. It implies a testable expectation: two tools should agree on where a network is weak even when they disagree on how weak, and a comparison designed around an identical route can isolate exactly this. The study therefore measures both networks along one shared path and classifies every sample by standard ranges so that the two tools can be compared category by category and location by location.

RESEARCH METHOD

Design, Site, and Instruments

This study uses a quantitative descriptive design built specifically for cross-instrument comparison. The essential design feature is that both measurement chains traversed an identical route at the same site, so that any difference in their output can be attributed to the instruments rather than to different measurement conditions. Measurements were taken at Cimory Dairyland, Bontoparang, Parangloe District, Gowa Regency, a hillside site enclosed by dense vegetation, along pedestrian visitor routes, between February and July 2025, for the networks of Telkomsel and XL Axiata. The first chain used Tems Pocket on an instrumented Android handset, with logfiles processed in TEMS Investigation; the second used Nemo Handy, with logfiles processed in Nemo Analyze. MapInfo prepared the route, and each chain logged RSRP, RSRQ, SINR, RSSI, and throughput continuously, geo-referencing each sample. The Tems Pocket unit was a mature device, a condition retained deliberately because it represents the realistic state of field equipment and allows the effect of instrument condition to be observed.

Reliability Comparison Procedure

The comparison was conducted at two levels for each parameter. At the level of absolute classification, the percentage of samples in each quality category was computed separately for each chain and the two figures were compared to quantify divergence. At the level of spatial pattern, the maps produced by each chain were inspected to determine whether they located the strongest and weakest areas in the same places. A parameter on which the chains agreed spatially but differed in absolute percentages was recorded as spatially reliable but not absolutely reliable; a parameter on which they agreed at both levels was recorded as fully reliable. The pattern of agreement and divergence across the five parameters was then generalised into a practical reliability rule for walk test reporting.

RESULTS AND DISCUSSION

Spatial Agreement Across Both Instruments

At the level of spatial pattern the two instruments agreed closely, as summarised in Table 1. For every parameter, both chains located the weakest values in the same southern and southwestern sectors of the park, the areas distant from the main building and screened by canopy and terrain, and both located the strongest values near the main building. For RSSI both chains reported abundant received power throughout the site, Telkomsel at 100 per cent strong on Tems Pocket and 91 per cent on Nemo Analyze; for SINR both placed the operators predominantly in the moderate band with the weakest points in the southwest; and for throughput both identified the same low-speed pockets. On the question of where the network is strong and weak, the two tools returned the same answer.

Table 1. Cross-Instrument Agreement by Parameter and Level

Parameter	Tems Pocket	Nemo Analyze	Spatial Agreement	Absolute Agreement
RSRP	80.34% poor (Telkomsel)	66.33% normal; 26.03% good	Agree (weakest in southwest)	Diverge sharply
RSRQ	XL Axiata mostly green	XL Axiata 54.43% yellow-orange	Agree	Diverge moderately
SINR	Moderate band dominant	61.20% at 10-20 dB (Telkomsel)	Agree	Broadly agree
RSSI	100% strong (Telkomsel)	91% above -70 dBm	Agree	Agree
Throughput	Same low-speed pockets	Same low-speed pockets	Agree	Broadly agree

Absolute Divergence, Most Sharply in RSRP

At the level of absolute classification, however, the two instruments diverged, and most sharply for RSRP. On the Tems Pocket chain 80.34 per cent of the Telkomsel route classified as poor to very poor, whereas on the Nemo Analyze chain the majority of the same route, 66.33 per cent, classified as normal with a further 26.03 per cent good. A reader given only the Tems Pocket figure would conclude the route was predominantly poor; a reader given only the Nemo Analyze figure would conclude it was predominantly adequate. The two conclusions are drawn from the same physical route and differ because of the instruments. RSRQ showed a similar though less extreme pattern, XL Axiata classifying favourably on Tems Pocket but falling to the yellow and orange classes on Nemo Analyze, while RSSI and the throughput classes agreed more closely.

The divergence admits a clear interpretation. Because it appears in absolute percentages while the spatial pattern is preserved, it is an instrument effect rather than a network effect: the two tools are observing the same physical variation but mapping it onto category thresholds differently, consistent with calibration differences and with the ageing of the Tems Pocket hardware. Had the divergence reflected a real network difference, it would have altered the spatial pattern too, placing the weak areas in different locations for the two tools; it did not. The parameter most affected, RSRP, is also the one whose absolute value is most sensitive to receiver calibration, which strengthens the instrument interpretation.

Interpreting the Divergence

The practical consequence is that the reliability of a walk test result depends on which level of the result is used. The spatial pattern is reliable: both instruments agree on it, so conclusions of the form "the southwestern sector is the weakest" are trustworthy from either tool alone. The absolute percentage is not reliable from a single instrument: the RSRP example shows that one tool can report a route as predominantly poor and another as predominantly normal. Any operational decision keyed to an absolute threshold, such as a contractual coverage percentage, therefore requires corroboration from a second tool, whereas a decision keyed to spatial priority, such as where to add a repeater, can rest on a single tool.

Synthesis: A Reliability Rule for Walk Testing

Generalising this pattern yields a reliability rule for walk test practice. First, report and act on spatial patterns, which are robust across instruments, in preference to absolute percentages, which are not. Second, where an absolute figure must be reported, corroborate it with a second, independently calibrated instrument before drawing conclusions from it. Third, treat instrument condition as a variable in its own right, disclosing the age and calibration state of the hardware, since an ageing unit can shift absolute values systematically. This rule does not diminish the value of walk testing; it specifies the level at which its results can be trusted, which is itself a contribution to measurement practice.

Discussion: Implications for Measurement Practice

The finding reframes how walk test results should be read and reported. The common practice of headlining an absolute percentage, a route is a given share in a given class, invests a single instrument with an authority the RSRP divergence shows it may not possess. Because two tools measuring the same route returned opposite verdicts on whether it was predominantly poor or predominantly normal, an absolute percentage from one tool is better understood as one instrument's reading than as a property of the network. The spatial pattern, by contrast, survived the change of instrument intact and can therefore carry the weight of operational conclusions. This is not an argument against walk testing but a specification of the level at which its output is dependable, and it aligns with the cautious use of multiple tools already visible in parts of the literature (Dewa et al., 2016; Herina, 2022).

Table 2. Reliability Rule and Practical Corollaries for Walk Testing

Result Level	Reliability	Recommended Practice	Rationale
Spatial pattern	Robust across tools	Report and act on it; single tool sufficient	Driven by physical propagation, common to both tools
Absolute percentage	Instrument-dependent	Corroborate with a second calibrated tool	RSRP verdict reversed between tools
Instrument condition	A variable in itself	Disclose hardware age and calibration state	Ageing units shift absolute values systematically
Operator ranking	Sensitive to tool choice	Base on corroborated, not single-tool, figures	A tool-dependent parameter could reverse the ranking

For practitioners, three implications follow directly. Studies that must report absolute figures should corroborate them across at least two independently calibrated instruments, and should disclose the condition of the hardware, since an ageing receiver can bias absolute values systematically as observed here. Studies whose purpose is to prioritise interventions, identifying where coverage is weakest, can rely on the spatial pattern from a single tool with confidence, because that level proved robust. And operator-comparison studies should be especially careful, because a parameter-level verdict that depends on the choice of instrument, as RSRP did here, could reverse an operator ranking if only one tool were used. Table 2 states the resulting reliability rule and its practical corollaries.

CONCLUSION

This study evaluated the cross-instrument reliability of Tems Pocket and Nemo Analyze by measuring the 4G LTE networks of Telkomsel and XL Axiata along an identical route at Cimory Dairyland and comparing their classifications of five key performance indicators. The two instruments agreed strongly on the spatial distribution of every parameter, both locating the weakest signal in the same vegetated southwestern sector, but diverged on absolute category percentages, most sharply for RSRP, where one chain classified 80.34 per cent of a route as poor while the other classified the majority as normal. Because the divergence appeared in absolute values while the spatial pattern was preserved, it is an instrument effect, consistent with calibration differences and the ageing of the Tems Pocket hardware, rather than a network effect.

The reliability rule that follows is that walk test conclusions are trustworthy at the level of spatial pattern but not at the level of absolute percentage from a single instrument. Spatial priorities, such as where to densify coverage, can rest on one tool; absolute figures, such as contractual coverage percentages, require corroboration from a second, independently calibrated tool, and the condition of the hardware should be disclosed. Future research should quantify the calibration offset between tools directly, extend the comparison to additional instruments and sites, and develop a standard reporting format that foregrounds spatial findings and flags absolute figures as instrument-dependent, so that the reliability of walk test evidence is transparent to those who act on it.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to the Department of Electrical Engineering, Ujung Pandang State Polytechnic, South Sulawesi, Indonesia, and the Department of Electrical Engineering, Universitas Kristen Indonesia Paulus, South Sulawesi, Indonesia, for their support, facilities, and valuable assistance throughout this research.

AUTHOR CONTRIBUTIONS

Conceptualization, S.M., A.T., and R.K.A.S; Methodology, S.M.; Investigation, A.T.; Data Collection and Formal Analysis, S.M., and R.K.A.S.; Writing Review & Editing, S.M., and A.T.,

CONFLICTS OF INTEREST

The author declares no conflict of interest.

REFERENCES

Dewa, I. G. A., Sudiarta, P. K., & Diafari, I. G. A. (2016). Analisis parameter jaringan HSDPA kondisi indoor dengan Tems Pocket Investigation dan G-NetTrack Pro. Jurnal Teknologi Elektro, 15(2), 45-52.

Efriyendro, R., & Rahayu, Y. (2017). Analisa perbandingan kuat sinyal 4G LTE antara operator Telkomsel dan XL Axiata berdasarkan parameter drive test menggunakan software G-NetTrack Pro. Jurnal Online Mahasiswa Fakultas Teknik Universitas Riau, 4(2), 1-9.

Fauzan, M., & Fery, A. (2018). Analisis performansi jaringan outdoor 4G LTE di kampus UMY antara provider Indosat dan Three. Jurnal Teknik Elektro Universitas Muhammadiyah Yogyakarta, 6(1), 33-42.

Herina, H. N. (2022). Analisis pengukuran kinerja jaringan 4G LTE berdasarkan hasil drive test. Politeknik Negeri Ujung Pandang.

Karo, F., Hikmaturokhman, A., & Danisya, A. R. (2019). Analisis hasil pengukuran performansi jaringan 4G LTE 1800 MHz di area Sokaraja Tengah Kota Purwokerto. Jurnal Infotel, 11(2), 55-63.

Krasniqi, F., Maraj, A., & Blaka, E. (2015). Performance analysis of mobile 4G/LTE networks. Proceedings of the South East European Conference on Telecommunications, 1-4.

Kurniawan, D., & Prasetyo, Y. (2020). Pengukuran kualitas sinyal jaringan LTE menggunakan walk test. Jurnal Telekomunikasi dan Jaringan, 8(1), 45-52.

Mursid, A. R., Hendrawan, H., & Sari, R. F. (2019). Analisis performansi dan optimisasi jaringan 4G LTE pada kawasan Telkom University. Jurnal Elektro dan Telekomunikasi Terapan, 6(1), 780-789.

Rahmania, R. (2020). Analisis cakupan area jaringan Long Term Evolution. Vertex Elektro, 12(10), 21-30

Sukiswo, S., & Zahra, A. A. (2018). Analisis quality of service (QoS) jaringan 4G dengan metode drive test pada kondisi outdoor. Jurnal Ilmiah Teknik Elektro, 7(2), 408-415.

Tamtama, B. (2017). Analisis kinerja kualitas jaringan 4G Long Term Evolution di kawasan Perumahan Singgalang, Kota Padang. Jurnal Teknik Elektro Universitas Negeri Padang, 5(2), 112-120.