

Journal of Applied Sciences, Information and Computing
Volume 6, Issue 2, Nov - Dec 2025
School of Mathematics and Computing, Kampala International University



ISSN: 3007-8903

<https://doi.org/10.59568/JASIC-2025-6-2-13>

Ultra-short scales in employee assessment: balancing efficiency and accuracy

E.O., Oladunmoye

Department of Applied Psychology, Kampala International University, Uganda.
oladunmoyetomenoch@gmail.com, <https://orcid.org/0000-0002-2721-2391>

Abstract

The increasing need for effective employee evaluation in the corporate world has stepped up application of ultra-short scales measures that have one to three items. These tools are useful because they save administration time, decrease survey fatigue, and reduce costs, although they pose serious psychometric challenges. The points presented in this paper are that ultra-short scales are at risk for construct underrepresentation, decreased reliability, and low validity, especially in regard to constructs with more than one dimension, such as burnout, engagement, and psychological safety. Empirical experiences prove that single-item and ultra-short measures are not effective to reflect the details needed to predict the outcomes because they can be effective only when the assessment targets the broad impressions. Their opponents note their practical advantages, such as higher response rates and popularity among managers in organizational environments where time is of the essence. In order to balance efficiency and rigor, this paper supports hybrid methods which make use of ultra-short scales as screening devices, supported by validated multi-item measures, and bolstered by predictive validation using artificial intelligence. The conclusion reiterates that ultra-short scales must not be substituted by the psychometric standards but rather used alongside them, so that efficiency, as well as accuracy, can be achieved in the making of organizational decisions.

Keywords: Ultra-short scales, Employee assessment, Psychometrics, Validity, Reliability, Survey efficiency

1. Introduction

The evaluation of employees has become a fundamental pillar of the current organizational strategy that allows enterprises to make evidence-based decisions in the areas related to recruitment, talent development, performance assessment, and employee well-being. Such tests give companies the information that can be used to shape recruitment activities, inform a professional development program, and establish possible hazards linked to stress, burnout, or disengagement (Richardson et al., 2025). Historically, psychometric tests employed in the field of organizational psychology have been

based on multi-item measures that could maximize reliability, address multidimensionality of constructs, and provide fairness across different groups of employees (Oz & Kanmaz, 2025).

In recent times, there has been an increasing trend in the tendency to utilize the short scales instruments that draw fewer items than the complete versions in order to shorten the time and cost expenses in the corporate environments. Scales that are ultra-short (containing one to three items per construct) are especially popular among them (Silva & Bortolon, 2025). They are more attractive because they do not need much of the

time of the employees, and their use minimizes survey fatigue and allows companies to obtain results on a vast sample. This effectiveness can be especially useful in intensive corporate settings, when extensive evaluations can be considered inefficient and response rates decline with the growth of a survey (Khan et al., 2024).

Despite these benefits, there is a critical issue of the psychometric quality of ultra-short scales. In theory, psychologically constructed variables, including burnout, engagement, or organizational commitment, are not simple constructs and would need a number of items to be properly captured by them (Kim et al., 2024). To reduce them to one item may result in the underrepresentation of the construct in question, and thus, compromised reliability and validity (Asli et al., 2024). Although ultra-short scales may be associated by a broad measure with longer scales, they may be inadequate to make refined differences that are helpful in diagnosing, predicting, and interfering (Smith, 2025).

Therefore, the rising popularity of ultra-short scales establishes a significant dilemma between efficiency and psychometric rigor. Organizations need to evaluate the time that they save by consuming fewer items against the possible loss of accuracy and reliability of assessment results. The main issue, however, is not whether ultra-short scales are convenient, but whether they can be valid replacements of longer and better-validated measures in the high-stakes environment of employee testing.

2. Methods

2.1 Risks of Construct Underrepresentation

In contrast to ultra-short scales which appear attractive in terms of their efficiency, the scales are very dangerous to the psychometric integrity of employee tests since it is exposed to construct underrepresentation. Sen (2024) explains that a tool is underrepresenting because it lacks the scope and the appropriateness to the construct it is supposed to measure, therefore, it provides a distorted or biased image. Multi-items, according to the classical test theory, are the most reliable and valid, as they allow sampling the construct domain more and minimizing the effect of measurement error (McClure et al., 2021). By contrast, ultra-short scales typically one to three items limit the representation of the construct, often neglecting important dimensions.

This weakness is especially tricky in the organizational setting where major constructs are multidimensional per se. As an example, burnout

is not an isolated phenomenon, but it consists of emotional fatigue, depersonalization, and decreased personal achievement (Vaishnav et al., 2025). On the same note, employee engagement comprises of vigor, dedication, and absorption (Maisyuri and Ariyanto, 2021), whereas psychological safety refers to interpersonal trust, openness, and risk-taking willingness at the workplace (Mitterer and Mitterer, 2023). The simplification of these complex constructs into a single or two items is dangerous as it might oversimplify the interactions amongst the components of the constructs.

Moreover, due to the lack of item variation, the scale is prone to measurement error and random response bias and, thus, reduced scale dependability (Oladunmoye, Oyedele, Sa'ad & Nakalema, 2024; Ali et al., 2021). The conclusions made by organizations such as calculating the risk of burnout, the level of engagement, or the workplace climate may be inaccurate in situations where the constructs are not captured. These inaccuracies can turn out to be costly as they can result in misplaced interventions, poor allocation of resources and mistrust of organizational surveys by the employees.

2.2 Empirical Findings on Ultra-Short Scales

There has been consistent empirical research on the trade-offs that are linked to ultra-short scales in assessment of workplace (Speer, Perrotta & Jacobs, 2024; Uhlig, Korunka, Prem & Kubicek, 2023). Although these measures enhance efficiency and minimise the respondent burden, they tend to compromise psychometric rigor especially on construct validity and reliability. An exemplary case study is the Copenhagen Burnout Inventory (CBI), which was originally designed as a multidimensional tool assessing personal, work-related, and client-related burnout. In some applied contexts, the CBI has been reduced to as few as three items.

Nevertheless, the literature shows that these reduced versions only take a limited perspective of the burnout experience, especially focusing on exhaustion but not on other critical aspects of burnout like depersonalization or diminished accomplishment (Reese, 2025; Hadzibajramovic et al., 2024; Schaufeli et al., 2020). This brings up the question of diagnostic accuracy of ultra-short burnout scales in determining levels of risk and interventions.

On the same note, studies that examine single-item indicators of job satisfaction and stress establish that in as far as these items are significant

predictors of their multi-item counterparts, their predictive validity is typically lower. The meta-analysis by Singh and Sant (2023) revealed that whereas these scales were found to be highly convergent when compared to the longer scales, they had a lower capacity to explain the critical outcome measures such as employee turnover, turnover, and job performance. This indicates that although single-item measurements might be suitable in a fast evaluation, they are not explanatory enough to make more complicated decisions about an organization. Similarly, Vinstrup et al., (2021) confirmed a single-item stress measure in research of occupational health, but they admitted that it was inadequate in distinguishing the physical, emotional and cognitive dimensions of stress. Organizations may end up over-simplifying themselves because of the use of such tools alone.

The limitations become even more apparent in the context of employee engagement. Schaufeli, Reisberger & Copuš (2025) argued that engagement is a multidimensional construct, consisting of vigor, dedication, and absorption. Compressing this construct into one or two items strips away these critical distinctions, thereby reducing the utility of the measure for targeted interventions. This indicates employee who reports high dedication but low vigor may require strategies aimed at reducing burnout and improving energy levels, information that would be lost in an ultra-short form measure. Thus, the inability of ultra-short scales to capture multidimensional complexity undermines their usefulness in guiding evidence-based organizational practices.

In addition to case-studies, meta-analytic data also supports these issues. The authors of the systematical review of short measures of personality traits and other psychological constructs, Lowman et al., (2025), found that the short versions are typically associated with lower reliability coefficients and weaker validity when compared to the full-length versions. Even though short measures have the benefit of usually increasing the response rate and survey completion the psychometrical trade-off is high in any survey research where the participation is of great importance (Schaufeli et al., 2020). Their conclusions point out that ultra-short scales can be handy in the exploratory method or in situations where the reduction of survey fatigue is the main priority, but they are much more inapplicable in high stakes scenarios like employee selection, promotion, or diagnostic evaluation.

2.3 Practical Advantages of Ultra-Short Scales

In addition to case-studies, meta-analytic data also supports these issues. The authors of the systematical review of short measures of personality traits and other psychological constructs, Lowman et al., (2025), found that the short versions are typically associated with lower reliability coefficients and weaker validity when compared to the full-length versions. Even though short measures have the benefit of usually increasing the response rate and survey completion the psychometrical trade-off is high in any survey research where the participation is of great importance. Their conclusions point out that ultra-short scales can be handy in the exploratory method or in situations where the reduction of survey fatigue is the main priority, but they are much more inapplicable in high stakes scenarios like employee selection, promotion, or diagnostic evaluation.

Another benefit is efficiency in fast-paced corporate environments, where decision-makers frequently need quick insights rather than lengthy diagnostic reports. In such cases, ultra-short scales provide a "snapshot" of employee attitudes, well-being, or engagement levels that can be used for timely decision-making (Fleten, 2025). During organizational change initiatives, leaders may prefer immediate feedback through one- or two-item indicators of stress or morale rather than administering comprehensive diagnostic tools. Importantly, research also suggests that for highly global constructs, such as *overall job satisfaction*, single-item measures may be sufficient. Pleeing (2022) and Guenther et al., (2025) demonstrated that single-item job satisfaction scales correlate strongly with multi-item measures, sometimes exceeding reliability expectations given their brevity (Fakunmoju, 2020). These findings indicate that while ultra-short scales may not be suitable for multidimensional constructs (e.g., engagement or burnout), they can perform adequately for broad, global constructs where precision in subdimensions is less critical.

2.4 Toward Hybrid and Technology-Enhanced Approaches

With these competing needs of efficiency and psychometric rigor, there is a middle ground to go in hybrid strategies on how to assess employees. Organizations can use ultra-short scales as a preliminary screening instrument, to be followed by more detailed assessments, when the results suggest the risks or areas of further interest. The

use of this tiered strategy will achieve efficiency in organizations and maintain diagnostic accuracy (Cao et al., 2025). An example of such is a one-item stress measure, which might be used as a quick check-in when conducting quarterly surveys, where employees report having a high degree of stress and then fill in a more detailed stress inventory.

The advances in artificial intelligence (AI) and machine learning are the other factors to suggest that hybrid assessment models that would guarantee efficiency and rigor can be achieved. The artificial intelligence systems can also combine ultra-short scale reactions with diverse organization information sources such as performance measures, absence rates, interactional patterns, and digital trace data to increase predictive validity and reduce the necessity to use bulky surveys (Salboux et al., 2025). By combining multiple streams of data, organizations will be capable of triangulating on several metrics of employee well-being, job satisfaction, or engagement scores, and such triangulation will be more precise and the psychometric tradeoffs of ultra-short instruments will be avoided by combining them. Not merely is this integration consistent with the trend of moving toward using more data-driven human resource management, but it also applies to decision-making (Elugbaju et al., 2024).

Computer-adaptive testing (CAT) is another line of thought that, along with AI incorporation, holds a promising road to brevity and psychometric validity reconciliation. With the assistance of algorithms, CAT dynamically chooses the next items, depending upon the previous answers of the respondent, and the organization can provide shorter but reliable tests (Weiss and Sahin, 2024; Oladunmoye & Oyedele, 2025a). As an example, employees whose responses are consistently alike can be asked only a few questions, whereas those with less predictable responses can be given some more questions in order to be more accurate. This customization helps in making evaluation more effective without compromising on measurement accuracy. Moreover, CAT approaches are consistent with fairness aspects because of minimized redundant respondent loads, which then ensure interest and a decrease in fatigue (Nobles et al., 2025; Oladunmoye et al., 2025b).

3. Conclusion

The benefits of ultra-short scales are undoubtedly practical in terms of employee evaluation,

particularly in those companies that are faced with survey fatigue, resource constraints, and urgent feedback requirements. They are efficient and scalable, which makes them appealing to the modern corporate environment. Nonetheless, all efficiency should not be realized at the cost of psychometric rigor. The evidence demonstrates that ultra-short scales are prone to construct underrepresentation, low reliability, and low validity especially with complex multidimensional constructs like burnout, engagement, and psychological safety.

The most rational solution is to use ultra-short scales to use as strategic preliminary predictors or screening instruments and rely on longer and validated scales and sophisticated analytics to get deeper information. The new technologies (AI and machine learning) allow more chances to balance the brevity and accuracy through the combination of the ultra-short responses with other data in the organization. Ultra-short scales are to be used as an addition and not a replacement to a known psychometric standard. Hybrid methods provide organizations with an opportunity to exploit efficiency and accuracy, and to make sure that the practice of assessment is scientifically sound and practically viable to facilitate high-stakes decisions.

4. References

- [1] Ali, A. M., Alkhamees, A. A., Hori, H., Kim, Y., & Kunugi, H. (2021). The depression anxiety stress scale 21: development and validation of the depression anxiety stress scale 8-item in psychiatric patients and the general public for easier mental health measurement in a post COVID-19 world. *International journal of environmental research and public health*, 18(19), 10142.
- [2] Asli, N. F., Matore, M. E. E. M., & Yunus, M. M. (2024). Construct validity of primary trait writing rubrics based on assessment use argument (AUA) validation framework. *Heliyon*, 10(22).
- [3] Cao, W., Mai, N. T., & Liu, W. (2025). Adaptive knowledge assessment via symmetric hierarchical Bayesian neural networks with graph symmetry-aware concept dependencies. *Symmetry*, 17(8), 1332.
- [4] Elugbaju, W. K., Okeke, N. I., & Alabi, O. A. (2024). Human resource analytics as a strategic tool for workforce planning and succession management. *International Journal of Engineering Research and Development*, 20(11), 744-756.

- [5] Fakunmoju, S. B. (2020). Validity of single-item versus multiple-item job satisfaction measures in predicting life: Satisfaction and turnover intention. *Asia-Pacific Journal of Management Research and Innovation*, 16(3), 210-228.
- [6] Fleten, M. K. (2025). *Great Expectations for AI: The Impact of Safety Climate and Strengths-Based Leadership on AI Literacy and Work Engagement among Norwegian Employees* (Master's thesis, NTNU).
- [7] Guenther, C., Lehnen, S., & Rilke, R. M. (2025). Formalization of annual performance feedback and employees' job satisfaction in the SME context. *Journal of Small Business Management*, 63(2), 558-589.
- [8] Hadžibajramović, E., Schaufeli, W., & De Witte, H. (2024). The ultra-short version of the Burnout Assessment Tool (BAT4)—development, validation, and measurement invariance across countries, age and gender. *Plos one*, 19(2), e0297843.
- [9] Khan, Y., Mustafa, S., Ahmad, R. W., Maqsood, T., Rehman, F., Ali, J., & Rodrigues, J. J. (2024). Content caching in mobile edge computing: a survey. *Cluster Computing*, 27(7), 8817-8864.
- [10] Kim, B. J., Jeon, N., Sohn, H., Lee, N., & Kim, M. J. (2024). The impact of corporate social responsibility on employee burnout: The crucial role of work overload. *Corporate Social Responsibility and Environmental Management*, 31(5), 4345-4360.
- [11] Lowman, K. L., Costello, T. H., Stewart, R. A., Joyner, K. J., Bertoldi, B. M., & Patrick, C. J. (2025). Creating Scale Measures of Latent Factors: A Genetic Algorithmic Approach. *Assessment*, 10731911251376230.
- [12] Maisyuri, R. S. I., & Ariyanto, E. (2021). The affect of vigor, dedication and absorption on the employee performance at Pt. Mandiri Utama Finance. *Dinasti International Journal of Management Science*, 2(4), 689-702.
- [13] McClure, K., Jacobucci, R., & Ammerman, B. A. (2021). Are items more than indicators? An examination of psychometric homogeneity, item-specific effects, and consequences for structural equation models.
- [14] Mitterer, D. M., & Mitterer, H. E. (2023). The mediating effect of trust on psychological safety and job satisfaction. *Journal of Behavioral and Applied Management*, 23(1), 29-41.
- [15] Nobles, C., Rangarajan, A., & Burrell, D. N. (2025). Human Factors Engineering-as-a-Service in Cybersecurity. In *New Horizons in Leadership: Inclusive Explorations in Health, Technology, and Education* (pp. 315-342). IGI Global Scientific Publishing.
- [16] Oladunmoye, E.O., and Oyedele, L.O., (2025a). Artificial Intelligence in Psychotherapy Implementation: Bridging Technology and Practice. *ISAR Journal of Medical and Pharmaceutical Sciences*, 3(5), 1-8.
- [17] Oladunmoye, E.O., Oyedele, O. Leah, Sa'ad, M.T., and Faith, Nakalema, (2024). Ethical Considerations in High-Stakes Academic Assessments: Insights from the Nigerian Educational System. *International Journal of Academic Multidisciplinary Research (IJAMR)*, 8(10), 101-108.
- [18] Öz, T., & Kanmaz, U. (2025). Structural modeling of factors influencing employee readiness for organizational change in aviation ground operations. *Öneri Dergisi*, 20(64), 205-238.
- [19] Pleeing, E. (2022). Measuring hope: Validity of short versions of four popular hope scales. *Quality & Quantity*, 56(6), 4437-4464.
- [20] Reese, J. (2025). *The Tipping Point: Executive Burnout, Brownout, and Realignment* (Doctoral dissertation, University of Pennsylvania).
- [21] Reisberger, P., & Copuš, L. (2025). Employee Engagement as a Driver of Satisfaction, Productivity, and Retention. In *Data-Centric Business and Applications* (pp. 311-321). Springer, Cham.
- [22] Richardson, E., Gordon, J., Ginnetti, R., Carroll, R., Cochran, R., Morris, L., ... & Brown, M. (2025). Employee Insight: Predicting, Visualizing, and Mitigating Intent to Leave. *Journal of Education Human Resources*, 43(2), 365-389.
- [23] Salbough, F., Mousavi, Y., Kucukdemiral, I. B., Fekih, A., & Cali, U. (2025). Reliability Assessment and Condition Monitoring of Wind Energy Conversion Systems Using Bayesian Networks: Recent Advances and Key Insights. *IEEE Access*.
- [24] Schaufeli, W. B., Desart, S., & De Witte, H. (2020). Burnout Assessment Tool (BAT)—development, validity, and reliability. *International journal of environmental research and public health*, 17(24), 9495.

- [25] Sen, I. (2024). *Identifying, characterizing, and mitigating errors in the automated measurement of social constructs from text data* (Doctoral dissertation, Dissertation, RWTH Aachen University, 2024).
- [26] Silva, E. M., & Bortolon, P. M. (2025). FK3BR Scale: Development of an Ultra-Short Measure.
- [27] Singh, S., & Sant, S. (2023). The moderating role of workplace (hybrid/remote) on employee engagement and employee turnover intention. *Employee responsibilities and rights journal*, 1-16.
- [28] Skogen, J. C., Andersen, A. I. O., Hjetland, G. J., Aarø, L. E., Colman, I., Sivertsen, B., & Finserås, T. R. (2025). Straightlining prevalence across domains of social media use and impact on internal consistency and mental health associations in the LifeOnSoMe study. *Scientific Reports*, 15(1), 28990.
- [29] Smith, T. W. (2025). Measurement and assessment in health psychology.
- [30] Speer, A. B., Perrotta, J., & Jacobs, R. R. (2024). Supervised construct scoring to reduce personality assessment length: A field study and introduction to the short 10. *Organizational Research Methods*, 27(2), 223-264.
- [31] Uhlig, L., Korunka, C., Prem, R., & Kubicek, B. (2023). A two-wave study on the effects of cognitive demands of flexible work on cognitive flexibility, work engagement and fatigue. *Applied Psychology*, 72(2), 625-646.
- [32] Vaishnav, V. A. D., Pandey, D. S., & Suyal, I. D. (2025). Structural Modelling of Burnout and Job Crafting In Start-Up Employees. *TPM–Testing, Psychometrics, Methodology in Applied Psychology*, 32(S3 (2025): Posted 07 July), 728-733.
- [33] Vinstrup, J., Jay, K., Jakobsen, M. D., & Andersen, L. L. (2021). Single-item measures of stress during work-and private time in healthcare workers. *Work*, 70(2), 583-589.
- [34] Weiss, D. J., & Sahin, A. (2024). *Computerized adaptive testing: From concept to implementation*. Guilford Publications.
- [35] Zhang, K., Zorba, D., Meda, A., Beerens, L., & Higham, D. J. Adversarial Attacks on Convolutional Neural Networks with General Purpose Optimization Software. *Modelling, Data Analytics and AI in Engineering*.