Art. #2608, 9 pages, https://doi.org/10.15700/saje.v46n1a2608

Mokken scale analysis of the Fourth Industrial Revolution teachers' effectiveness scale using automated selection procedure

Deborah Chidubem Adamu
Department of Education Leadership and Management, Faculty of Education, University of Johannesburg, Johannesburg, South Africa
adamuchidubemdeborah@gmail.com

Abstract
The Fourth Industrial Revolution (4IR) has led to the need for a redefinition of teacher effectiveness, particularly in science, technology, engineering, and mathematics (STEM) education areas such as chemistry. In the study reported on here I used a quantitative, exploratory cross-sectional survey design to validate the 4IR chemistry teachers' effectiveness scale (4IRCTES) using Mokken scale analysis (MSA) with the automated item selection procedure (AISP). The study population included chemistry teachers and senior secondary school students from public and private schools in Osun and Oyo States, Nigeria. Data were collected from 35 teachers and 320 students through multistage and proportionate stratified random sampling. The 4IRCTES, originally consisting of 49 items across 6 domains, underwent expert content validation. A pilot study involving 20 teachers and 50 students resulted in a Cronbach's alpha of 0.87, indicating strong internal consistency. The AISP analysis revealed 5 interpretable sub-scales, each representing a unidimensional construct, confirming the multidimensional structure of the 4IRCTES and supporting its construct validity and psychometric reliability. The findings show that the 4IRCTES is a reliable and valid tool for evaluating chemistry teacher effectiveness in technology-driven educational settings. This study emphasises the value of nonparametric item response theory methods like MSA in the validation of educational instruments and highlights the essential 4IR competencies needed for effective STEM teaching in Nigerian secondary schools.

Keywords: automated item selection procedure; chemistry education; fourth industrial revolution; Mokken scale analysis; Nigerian secondary schools; nonparametric item response theory (IRT); psychometric validation; teacher effectiveness

Introduction
The Fourth Industrial Revolution (4IR) has significantly changed the education landscape, especially in the field of science education. With rapid advancements in artificial intelligence, robotics, big data, and digital technologies, the 4IR requires a re-evaluation of teaching effectiveness to include skills like computational thinking, data literacy, and technology-enhanced pedagogy (Aboderin & Havenga, 2024; Schwab, 2017; Xu, David & Kim, 2018). In this context, chemistry education, crucial for STEM advancement, needs a nuanced approach to teacher evaluation that incorporates traditional teaching skills and emerging digital-age competencies (Eilks & Hofstein, 2017; United Nations Educational, Scientific and Cultural Organization [UNESCO], 2021).

In Nigeria, where chemistry is a fundamental subject in secondary school STEM curricula, there is a pressing need to empower teachers with the necessary skills to prepare students for the digital economy (Federal Ministry of Education, 2018). However, existing teacher evaluation tools often overlook 4IR competencies, limiting their effectiveness in today's educational landscape (Organisation for Economic Co-operation and Development [OECD], 2019). To bridge this gap, the Fourth Industrial Revolution chemistry teachers' effectiveness scale (4IRCTES) was developed as a context-specific tool to assess chemistry teachers' effectiveness in both traditional teaching methods and 4IR-relevant skills (Adamu, 2023).

Validating the 4IRCTES is crucial in ensuring its reliability and practicality in real-world educational settings. In this study I used Mokken scale analysis (MSA), a nonparametric item response theory (IRT) technique, to investigate the dimensionality and psychometric properties of the 4IRCTES. MSA is suited well for educational settings in developing countries where response data may not meet the assumptions of parametric models (Ayanwale & Ndlovu, 2021; Sijtsma & Van der Ark, 2017). I specifically employed the automated item selection procedure (AISP) within MSA to analyse the scale's latent structure, identify poorly performing items, and determine the number of meaningful sub-scales.

However, empirical validation of the 4IRCTES is still lacking, especially when using psychometric approaches that are suitable for nonparametric educational data. Moreover, limited research exists on teacher effectiveness assessment tools that explicitly capture 4IR competencies in African secondary education systems. These gaps necessitated a rigorous validation of the 4IRCTES using appropriate item response models such as MSA. The purpose of this study was to introduce the AISP as a framework for determining both the number of underlying factors and the interpretable factor structure of the 4IRCTES through MSA. The specific research objectives were:

1) To determine the number of factors underlying the 4IRCTES by applying MSA scalability criteria.
2) To identify interpretable factors that demonstrate adequate psychometric properties for measuring chemistry teachers' effectiveness in 4IR contexts.

I addressed these objectives using the following research questions:

1) What is the number of factors underlying the 4IRCTES as determined by MSA?
2) How many interpretable factors with adequate scalability coefficients underlie the 4IRCTES?

Literature Review

Conceptualising teacher effectiveness in the 4IR era

Teacher effectiveness has traditionally been defined in terms of instructional clarity, classroom management, subject knowledge, and student outcomes (Stronge, 2018). However, the pedagogical demands of the 4IR have expanded this definition to include the ability to incorporate digital tools, computational thinking, problem-solving, and adaptive instruction into classroom practice (Chaka, 2020; OECD, 2019). These emerging competencies are especially critical in subjects such as chemistry, where hands-on experimentation and data interpretation can be significantly enhanced through technological integration (Kelly, 2020).

The 4IRCTES was developed in response to these evolving educational needs. The instrument includes items that measure competencies in traditional teaching, the use of technology in the classroom, data-driven instruction, digital content creation, and student engagement in technology-rich environments (Adamu, 2023). Despite the theoretical relevance of the 4IRCTES, empirical validation was necessary to establish its psychometric soundness and interpretability.

Mokken scale analysis as a tool for validation

MSA was introduced by Mokken (1971) as a probabilistic extension of Guttman's deterministic scalogram model (Guttman, 1944). It offers a robust framework for examining the scalability and unidimensionality of ordinal items in psychological and educational instruments. Unlike parametric IRT models that require assumptions about the shape of item response functions (IRF) and population distributions (Embretson & Reise, 2013), MSA is nonparametric and, therefore, more flexible in analysing Likert-type data (Sijtsma & Van der Ark, 2017).

MSA consists of two hierarchical models: the monotone homogeneity model (MHM) and the double monotonicity model (DMM). The MHM requires three assumptions – unidimensionality, monotonicity, and local independence, while the DMM adds the more restrictive assumption of invariant item ordering (Ligtvoet, Van der Ark, Te Marvelde & Sijtsma, 2010; Van der Ark, 2007). These models allow researchers to evaluate whether a set of items measures a single latent trait and whether items maintain consistent ordering across respondent ability levels.

The use of the AISP within MSA has further enhanced its utility in scale development and validation. AISP applies an algorithmic approach to form unidimensional item clusters (scales) based on a minimum item scalability threshold (Hi $\geq$ 0.3 or 0.4), thereby identifying both strong and weak items (Meijer & Baneke, 2004; Straat, Van der Ark & Sijtsma, 2013). This objective method complements traditional factor analysis by offering fine-grained insight into the psychometric structure of instruments.

Applications of MSA in educational research

MSA has gained increasing recognition for its practical advantages in educational research, especially in contexts with limited statistical resources or ordinal data structures. Tabatabaee-Yazdi, Samir, Bakhtshirin and Tabatabaeiyazdi (2024) have, for instance, developed and validated a 36-item inventory to assess mobile teaching affordances among 204 teachers of English as a foreign language (EFL). Gao, Bai, Sun and Jia (2022) developed and validated a 14-item career preference questionnaire for Chinese medical undergraduates. MSA was used to refine the questionnaire, resulting in two unidimensional sub-scales: a 10-item career advantage sub-scale and a 4-item career disadvantage sub-scale. The final instrument demonstrated acceptable reliability and construct validity, making it a suitable tool for exploring medical students' career motivations. Ayanwale and Ndlovu (2021) used MSA to evaluate the scalability of mathematics items in West African examinations and found that over 20% of the items exhibited poor psychometric properties. Palmgren, Brodin, Nilsson, Watson and Stenfors (2018) applied MSA to the Dundee ready educational environment measure (DREEM) to assess its psychometric properties among 222 undergraduate physiotherapy students.

Despite these developments, significant gaps remain. Firstly, the application of MSA in teacher effectiveness research, particularly for 4IR-specific constructs – remains limited (Wind, 2017). Secondly, advanced MSA techniques like AISP are underutilised in African contexts, despite their suitability for emerging educational assessment needs (Meijer & Sijtsma, 2001; Van der Ark, 2007). Lastly, there is a dearth of validated instruments specifically designed to assess 4IR-related competencies of secondary school chemistry teachers in Nigeria (Teo, Unwin, Scherer & Gardiner, 2021). I used this study to address these gaps and introduce the 4IRCTES as a novel tool for measuring chemistry teachers' effectiveness in the 4IR era and to apply MSA, through AISP, to evaluate its psychometric structure and dimensionality.

Methodology

Research Design

I used a quantitative, exploratory cross-sectional survey design to investigate the psychometric properties of the 4IRCTES using MSA. The selection of a quantitative design was appropriate due to its ability to gather measurable data from a large sample to analyse patterns and structures underlying psychometric constructs (Creswell & Creswell, 2018). The exploratory nature of the design aligns with the objective of uncovering latent constructs using a data-driven approach like MSA (Van Schuur, 2003).

Population and Sample

The study population consisted of public and private senior secondary school chemistry teachers and students in Osun and Oyo States, Nigeria. These states were purposefully chosen due to their diverse educational settings and representation of secondary school contexts in southwestern Nigeria. A multistage sampling technique was used. Firstly, two states (Osun and Oyo) were selected. Secondly, three local government areas (LGAs) from each state were randomly chosen. Lastly, through proportionate stratified random sampling, a total of 35 chemistry teachers and 320 senior secondary school chemistry students (SSS1–SSS3) were selected from the designated schools. The sample size was deemed sufficient for nonparametric IRT analysis. According to Meijer and Baneke (2004), a sample size of at least 300 participants is suitable for reliable MSA. Additionally, including teachers and students offers a multi-informant perspective, enhancing the validity of the results.

Instrumentation

The 4IRCTES was self-designed to encompass various dimensions of teachers' effectiveness aligned with the requirements of 4IR teaching environments. Through a review of literature on teachers' effectiveness and the specific needs of Nigerian secondary schools, six key dimensions were identified to guide item development:

1) Teachers' personality
2) Classroom management and organisation
3) Organising and orienting for instruction
4) Implementing instruction
5) Monitoring students' progress and potential
6) Professional development

I employed a structured rating instrument, 4IRCTES, in the study. Initially consisting of 49 items, I developed the scale to evaluate effectiveness across different 4IR teaching domains in chemistry education. Each item was rated on a 4-point Likert scale: 1 (Very poor), 2 (Poor), 3 (Moderate), and 4 (Good). The initial item pool underwent expert review for content validity by university scholars in science education and educational measurement, as suggested by Boateng, Neilands, Frongillo, Melgar-Quiñonez and Young (2018). The instrument was pilot tested with 20 chemistry teachers and 50 senior students from secondary schools outside the main study area to assess reliability and clarity. The preliminary reliability analysis indicated acceptable internal consistency with a Cronbach's alpha of 0.87 (Tavakol & Dennick, 2011).

Data Collection Procedure

Data collection took place over a 4-week period. Prior to administering the data, official permission was obtained from school principals, and participants provided informed consent. Trained assistants and I distributed the questionnaire in person, ensuring that consistent instructions were provided and ethical standards were maintained. Participants were assured of their anonymity and confidentiality in accordance with the University of Johannesburg's (UJ) research ethics policy (UJ, 2021).

Teachers were instructed to independently complete the 4IRCTES rating scale, while students completed a parallel version rating their chemistry teachers. The dual-rating method aided in triangulating responses; a practice recommended in scale development for enhanced validity (DeVellis, 2017).

Data Analysis

Data were analysed using the AISP within the MSA framework, a nonparametric item response theory (NIRT) model. MSA is particularly beneficial for exploring latent hierarchical structures in questionnaire data without assuming normal distribution, making it suitable for validating educational and psychological scales (Sijtsma & Van der Ark, 2017). MSA is advantageous in verifying the assumption of item response model unidimensionality and empirical monotonicity (Sijtsma & Molenaar, 2002). MSA also enables the identification of scalable sub-dimensions within a scale, crucial for refining instruments for multidimensional constructs like teacher effectiveness (Van Schuur, 2003).

Although MSA typically involves several analytical steps, I specifically focused on applying AISP to determine the number of underlying factors and their interpretability in alignment with the study objectives. Below are the standard steps in MSA, followed by a description of those adopted in this study.

Standard steps in Mokken scale analysis

1) Data screening: Screening for missing values and outliers was conducted, and cases with more than 10% missing data were excluded.
2) Computation of item-pair scalability coefficients (Hij): Evaluates relationships between item pairs.

3) Computation of item scalability coefficients (Hi): Assesses each item's strength in measuring the latent construct.
4) Computation of overall scale scalability coefficient (H): Evaluates the overall scale strength.
5) AISP: The AISP algorithm was used to group items into scalable sub-scales (Van der Ark, 2007). Items that did not meet scalability thresholds or formed part of factors with fewer than three items, were removed.
6) Factor interpretability: Sub-scales were examined for conceptual coherence by interpreting the underlying themes of grouped items. Sub-scales with only two items were discarded, as two-item factors are not considered stable for psychological constructs (Worthington & Whittaker, 2006).

Aligned with the research objectives, I specifically focused on Steps 1, 5, and 6. The data were screened for completeness, and the AISP algorithm (Van der Ark, 2007) was applied to determine the number of latent factors underlying the 4IRCTES. Items forming interpretable and thematically coherent factors (with three or more items) were retained, while those in sub-scales with fewer than three items were excluded, consistent with psychometric standards (Worthington & Whittaker, 2006). Steps involving the computation of individual and overall scalability coefficients (Hi, Hij, and H) were acknowledged but not applied, as they were beyond the specific scope of this analysis.

Ethical Considerations

Ethical approval was obtained from the Research Ethics Committee of the Faculty of Education at the Obafemi Awolowo University, Nigeria. Participants were informed of their right to withdraw from the study without penalty at any time. Data were anonymised and securely stored. The study adhered to the Declaration of Helsinki's ethical principles for human subject research (World Medical Association, 2013).

Results

The results obtained are based on the research questions presented in this study. The analysis was conducted using the AISP, an exploration model of MSA, to determine the underlying factor structure of the 4IRCTES.

First Research Question: What is the Number of Factors Underlying the 4IRCTES as Determined by MSA?

To address this research question, an exploratory MSA was conducted using the AISP to identify scalable sub-scales of the 4IRCTES. The goal was to determine whether multiple sub-scales could yield a stronger measure of 4IRCTES compared to a single-factor model. Table 1 shows the factors and items identified by AISP as scalable, including their scalability coefficients.

Table 1 Initial scalable factors identified by AISP for the 4IRCTES (49 items, 13 factors)

Factor	Serial number	Item	Scalability coefficient	Factor	Serial number	Item	Scalability coefficient
1	1	IT1	0.65	4	26	IT35	0.49
	2	IT2	0.65		27	IT36	0.49
	3	IT3	0.58	5	28	IT13	0.49
	4	IT7	0.53		29	IT14	0.49
	5	IT19	0.5		30	IT10	0.46
	6	IT9	0.49		31	IT11	0.45
	7	IT17	0.48	6	32	IT38	0.49
	8	IT32	0.46		33	IT39	0.49
	9	IT33	0.46		34	IT37	0.45
	10	IT31	0.45	7	35	IT26	0.47
	11	IT30	0.45		36	IT27	0.47
	12	IT34	0.45	8	37	IT61	0.45
	13	IT25	0.45		38	IT62	0.45
	14	IT28	0.45	9	39	IT15	0.44
	15	IT29	0.44		40	IT16	0.44
	16	IT12	0.44	10	41	IT54	0.44
2	17	IT20	0.6		42	IT55	0.44
	18	IT22	0.6	11	43	IT45	0.44
	19	IT23	0.56		44	IT46	0.44
3	20	IT48	0.51	12	45	IT5	0.43
	21	IT49	0.51		46	IT6	0.43
	22	IT50	0.47	13	47	IT41	0.42
	23	IT51	0.46		48	IT43	0.42
	24	IT52	0.45		49	IT42	0.42
	25	IT53	0.45				

Table 1 indicates that initially, 13 scalable sub-factors consisting of 49 items were identified. However, seven of the 13 factors (factors 4, 7, 8, 9, 10, 11, and 12) were found to have fewer than three items each, making them statistically uninterpretable. For example, factor 4 had items 35

and 36, factor 7 had items 26 and 27, factor 8 had items 61 and 62, factor 9 had items 15 and 16, factor 10 had items 54 and 55, factor 11 had items 45 and 46, and factor 12 had items 5 and 6 loaded on them. These seven sub-factors loaded only two items each, totalling 14 items. Therefore, the seven sub-scales, with 14 items in total, were removed from further analysis. As a result, only six interpretable and scalable sub-scales (factors 1, 2, 3, 5, 6, and 13) were retained, representing a total of 35 items (that is, 49 less 14). These factors were renamed sequentially (1 to 6) for ease of interpretation.

Second Research Question: How Many Interpretable Factors with Adequate Scalability Coefficients Underlie the 4IRCTES?

The six interpretable factors with the items loading on them are shown in Table 2. Each factor was interpreted based on the conceptual similarity among its items.

Table 2 Final interpretable factors and items of the 4IRCTES (six factors, 35 items)

Factor	Serial number	Item	Scalability coefficient
1	1	IT1	0.65
	2	IT2	0.65
	3	IT3	0.58
	4	IT7	0.53
	5	IT19	0.5
	6	IT9	0.49
	7	IT17	0.48
	8	IT32	0.46
	9	IT33	0.46
	10	IT31	0.45
	11	IT30	0.45
	12	IT34	0.45
	13	IT25	0.45
	14	IT28	0.45
	15	IT29	0.44
	16	IT12	0.44
2	17	IT20	0.6
	18	IT22	0.6
	19	IT23	0.56
3	20	IT48	0.51
	21	IT49	0.51
	22	IT50	0.47
	23	IT51	0.46
	24	IT52	0.45
	25	IT53	0.45
4	26	IT13	0.49
	27	IT14	0.49
	28	IT10	0.46
	29	IT11	0.45
5	30	IT38	0.49
	31	IT39	0.49
	32	IT37	0.45
6	33	IT41	0.42
	34	IT43	0.42
	35	IT42	0.42

Upon closer analysis, it was discovered that items in Factor 6 (IT41, IT43, and IT42) aligned with two separate constructs – implementing instruction and monitoring students' progress and potential – rather than forming a cohesive standalone factor. Therefore, Factor 6 and its three items were removed from the final structure.

As a result, the final version of the 4IRCTES consisted of five clear factors and 32 items, arranged as follows:

- Teachers' personality: IT1, IT2, IT3, IT7
- Classroom management and organisation: IT19, IT9, IT17, IT12, IT20, IT22, IT23, IT13, IT14, IT10, IT11
- Organising and orienting for instruction: IT30, IT25, IT28, IT29
- Implementing instruction: IT32, IT33, IT31, IT34, IT38, IT39, IT37
- Digital and professional competencies: IT48, IT49, IT50, IT51, IT52, IT53

While the scale includes a distinct sub-scale, digital and professional competencies, its content closely aligns with the broader definition of professional development in the study. It demonstrates how teachers use lifelong learning and training to navigate technological advancements in the classroom, making it an applied form of professional development within the 4IR

framework. With the study I confirmed that the 4IRCTES is a multidimensional tool, defined by five key constructs and 32 items, which enhances its interpretability and structural coherence.

Discussion

In this study I used MSA with the AISP to examine the psychometric properties and factorial structure of the 4IRCTES. This approach, based on nonparametric IRT, allowed for the identification of hierarchical, interpretable sub-scales that represent latent traits associated with 4IR teacher effectiveness in Nigerian secondary education.

The analysis revealed five unidimensional and interpretable sub-scales: (1) teachers' personality, (2) classroom organisation and management, (3) implementing instruction, (4) monitoring and evaluation of students, and (5) digital and professional competencies. These factors align with global concepts of effective teaching in technologically advancing classrooms (Danielson, 2007; Stronge, 2018).

The emergence of a digital and professional competencies sub-scale is of theoretical significance, highlighting how the 4IR has expanded teaching effectiveness to include not only pedagogical skills but also proficiency in emerging digital tools, data literacy, and artificial intelligence (AI) integration (OECD, 2021; Schwab & Davis, 2018). This validates earlier claims by Ajani and Govender (2023), and Okunlola (2024) that modern teacher effectiveness frameworks must incorporate digital proficiency to meet contemporary classroom needs.

The identification (that is, unidimensionalty) of the five sub-scales confirms the multidimensional nature of the 4IRCTES, supporting its potential use in various evaluation and professional development contexts. The grouping of items into coherent and interpretable domains showcases the ability of MSA to reveal latent structures without the constraints of parametric IRT. This aligns with previous recommendations for adaptable, context-sensitive tools suitable for developing country contexts (Sijtsma & Van der Ark, 2017). This result also aligns with research by Ajani and Govender (2023); Malunda, Onen, Musaazi and Oonyu (2016); and Saka and Onanuga (2019), emphasising the necessity for teacher evaluation tools to address domain-specific challenges, especially in science education and resource-constrained environments.

Furthermore, the success of the AISP in clustering items into hierarchical and reliable scales without imposing parametric assumptions on items confirms the suitability of Mokken scaling for instrument development in developing country contexts. Traditional parametric IRT methods require larger sample sizes and normality assumptions, which may not always be feasible in African educational settings (Sijtsma & Van der Ark, 2017). Therefore, with this study I methodologically demonstrate the applicability of MSA in educational measurement research among underrepresented populations (Abdelhamid, Gómez-Benito, Abdeltawwab, Abu Bakr & Kazem 2020; Ayanwale & Ndlovu, 2021; Tabatabaee-Yazdi et al., 2024). Another significant implication is the inclusion of student perspectives in assessing teacher effectiveness. By analysing students' feedback on their chemistry teachers, I support the idea that students' viewpoints offer valuable insight into classroom practices, particularly those involving digital tools and 4IR teaching methods (Fauth, Decristan, Rieser, Klieme & Büttner, 2014; Peterson, Wahlquist & Bone, 2000; Stronge, 2018). This triangulation approach enhances the instrument's validity and supports its use in formative assessment and school improvement.

The results from the AISP in MSA strongly endorse the scalability and dimensionality of the 4IRCTES (Mokken, 1971). The five interpretable scales that emerged align with the theoretical foundations of effective 4IR teaching practices. The robust scalability coefficients (Hi) across the items indicate a high level of internal consistency and item discrimination within each scale (Sijtsma & Molenaar, 2002). Overall, the findings suggest that the 4IRCTES is a valid tool for evaluating chemistry teachers' effectiveness within the 4IR context, with potential for broader application in science teaching domains in sub-Saharan Africa.

Conclusion

With this study I provided a psychometric validation of the 4IRCTES using MSA with AISP. The findings confirm that the scale demonstrates acceptable internal consistency, scalability, and construct validity, resulting in five interpretable and meaningful sub-scales.

By employing a nonparametric IRT framework, the study contributes to the development of context-specific and statistically robust instruments tailored to the realities of Nigerian secondary schools (Sijtsma & Molenaar, 2002). The structure of the 4IRCTES reflects the changing landscape of science education, where teacher effectiveness is increasingly defined by adaptability, digital fluency, student engagement, and evidence-based practice (Darling-Hammond, Flook, Cook-Harvey, Barron & Osher, 2020; Schleicher, 2018).

In summary, this study offers a theoretically grounded and empirically validated scale for evaluating the competencies required of chemistry teachers in the digital age. It fills a crucial gap in educational measurement and supports the professionalisation and accountability of teachers in emerging economies.

Implications of the Study

The study has theoretical, methodological and practical implications. Firstly, it expands the literature on teacher effectiveness measurement by validating a culturally and technologically relevant instrument rooted in the realities of 4IR education in Africa. It supports the view that effective teaching is multidimensional, integrating personality, pedagogy, monitoring, and digital competence (Danielson, 2007; Stronge, 2018).

Secondly, the successful application of MSA confirms the value of nonparametric IRT in contexts where sample sizes and data characteristics may limit traditional methods. It provides a framework for instrument development in developing countries using AISP and MSA, which can be replicated by educational researchers working in similar low-resource environments.

Lastly, the validated 4IRCTES can serve as a diagnostic tool for school leaders, curriculum developers, and policymakers to identify strengths and gaps in teacher performance, particularly in science education. The study offers a foundation for professional development programmes, helping education ministries and teacher training institutions tailor interventions to specific competencies (e.g., digital pedagogy, classroom management). The tool can be adapted to other STEM subjects, broadening its utility and informing nationwide teacher appraisal systems aligned with the 4IR agenda.

Limitations of the Study

Despite the promising findings of this study, several limitations are acknowledged. Firstly, the study was geographically limited to two states in southwestern Nigeria (Osun and Oyo), which may affect the generalisability of the findings to other regions of the country or broader African contexts. Differences in infrastructure, teacher training, and technology integration across regions may influence the applicability of the 4IRCTES in other settings.

Secondly, while MSA is well-suited for ordinal data and non-normal distributions, it does not provide item-level parameter estimates like parametric IRT models. This limits the ability to explore finer-grained item characteristics such as item discrimination and difficulty, which may be necessary for refining the scale further.

Thirdly, the sample size, although adequate for MSA, may not be sufficiently large for more complex analyses such as confirmatory factor analysis or multi-group invariance testing. These additional analyses would provide deeper insight into the stability and robustness of the scale across diverse teacher and student populations.

Fourthly, although the 4IRCTES was pilot-tested and reviewed by experts for content validity, further validation using convergent and discriminant validity measures was beyond the scope of this study. Such validity checks, along with criterion-related validity studies, would enhance confidence in the instrument's overall utility and interpretability.

Lastly, in this study I relied primarily on self-reported and observer-rated data. While these methods are valuable, they may be influenced by social desirability bias or subjective perceptions. Future studies could incorporate classroom observations, student achievement data, or digital teaching portfolios to triangulate findings and validate teacher effectiveness from multiple data sources.

Suggestions for Further Studies

1) Future research should replicate this study in other geopolitical zones of Nigeria or other African countries to test the cross-cultural validity and generalisability of the 4IRCTES.
2) Researchers are encouraged to conduct confirmatory factor analysis (CFA) and IRT modelling to complement MSA findings and assess item-level properties.
3) Future studies should explore the predictive validity of the 4IRCTES by examining its relationship with student learning outcomes and teacher professional development engagement.
4) Longitudinal research design could be adopted to assess how teachers' effectiveness in 4IR competencies evolves over time, especially after targeted training interventions.
5) Qualitative follow-up studies, such as interviews or focus groups with teachers, could offer richer insight into the contextual factors that influence 4IR teaching effectiveness.

Conflict of Interest

The author does not have any conflict of interest to declare.

Notes

ii. DATES: Received: 4 June 2024; Revised: 7 May 2025; Accepted: 4 February 2026; Published: 28 February 2026.

References

Abdelhamid GSM, Gómez-Benito J, Abdeltawwab ATM, Abu Bakr MHS & Kazem AM 2020. A demonstration of Mokken scale analysis methods applied to cognitive test validation using the Egyptian WAIS-IV. *Journal of Psychoeducational Assessment*, 38(4):493–506. https://doi.org/10.1177/0734282919862144

Aboderin OS & Havenga M 2024. Essential skills and strategies in higher education for the fourth industrial revolution: A systematic literature review. *South African Journal of Higher Education*, 38(2):24–43. https://doi.org/10.20853/38-2-5430

Adamu CD 2023. Dimensionality of chemistry teachers' effectiveness scale (CTES) in secondary schools in

Osun state, Nigeria. *International Journal of Studies in Psychology*, 3(2):77–81. https://doi.org/10.38140/ijspsy.v3i2.941

Ajani OA & Govender S 2023. Impact of ICT-driven teacher professional development for the enhancement of classroom practices in South Africa: A systematic review of literature. *Journal of Educational and Social Research*, 13(5):116–128. https://doi.org/10.36941/jesr-2023-0125

Ayanwale MA & Ndlovu M 2021. Ensuring scalability of a cognitive multiple-choice test through the Mokken package in R programming language. *Education Sciences*, 11(12):794. https://doi.org/10.3390/educsci11120794

Boateng GO, Neilands TB, Frongillo EA, Melgar-Quiñonez HR & Young SL 2018. Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, 6:149. https://doi.org/10.3389/fpubh.2018.00149

Chaka JG & Govender I 2020. Implementation of mobile learning using a social network platform: Facebook. *Problems of Education in the 21st Century,* 78(1):24–47. https://doi.org/10.33225/pec/20.78.24

Creswell JW & Creswell JD 2018. *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed). Thousand Oaks, CA: Sage.

Danielson C 2007. *Enhancing professional practice: A framework for teaching* (2nd ed). Alexandria, VA: Association for Supervision and Curriculum Development.

Darling-Hammond L, Flook L, Cook-Harvey C, Barron B & Osher D 2020. Implications for educational practice of the science of learning and development. *Applied Developmental Science*, 24(2):97–140. https://doi.org/10.1080/10888691.2018.1537791

DeVellis RF 2017. *Scale development: Theory and applications* (4th ed). Thousand Oaks, CA: Sage.

Eilks I & Hofstein A 2017. Curriculum development in science education. In KS Taber & B Akpan (eds). *Science education: An international course companion*. Rotterdam, The Netherlands: Sense. https://doi.org/10.1007/978-94-6300-749-8_13

Embretson SE & Reise SP 2013. *Item response theory for psychologists*. New York, NY: Psychology Press.

Fauth B, Decristan J, Rieser S, Klieme E & Büttner G 2014. Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction*, 29:1–9. https://doi.org/10.1016/j.learninstruc.2013.07.001

Federal Ministry of Education 2018. *National policy on science and technology education*. Abuja, Nigeria: Author. Available at https://education.gov.ng/wp-content/uploads/2020/09/National-Policy-On-Science-and-Technology-Education.pdf. Accessed 10 May 2025.

Gao Y, Bai X, Sun L & Jia D 2022. Development of a career questionnaire for medical undergraduates using Mokken scale analysis. *BMC Medical Education*, 22:286. https://doi.org/10.1186/s12909-022-03340-8

Guttman L 1944. A basis for scaling qualitative data. *American Sociological Review*, 9(2):139–150. https://doi.org/10.2307/2086306

Kelly EW 2020. Reflections on three different high school chemistry lab formats during COVID-19 remote learning. *Journal of Chemical Education*, 97(9):2606−2616. https://doi.org/10.1021/acs.jchemed.0c00814

Ligtvoet R, Van der Ark LA, Te Marvelde JM & Sijtsma K 2010. Investigating an invariant item ordering for polytomously scored items. *Educational and Psychological Measurement*, 70(4):578–595. https://doi.org/10.1177/0013164409355697

Malunda P, Onen D, Musaazi JCS & Oonyu J 2016. Teacher evaluation and quality of pedagogical practices. *International Journal of Learning, Teaching and Educational Research*, 15(9):18–133. Available at https://nru.uncst.go.ug/handle/123456789/5094. Accessed 10 May 2025.

Meijer RR & Baneke JJ 2004. Analyzing psychopathology items: A case for nonparametric item response theory modeling. *Psychological Methods*, 9(3):354–368. https://doi.org/10.1037/1082-989X.9.3.354

Meijer RR & Sijtsma K 2001. Methodology review: Evaluating person fit. *Applied Psychological Measurement*, 25(2):107–135. https://doi.org/10.1177/01466210122031957

Mokken RJ 1971. *A theory and procedure of scale analysis: With applications in political research*. Berlin, Germany: De Gruyter Mouton. https://doi.org/10.1515/9783110813203

Okunlola JO 2024. Digital technology adoption and school leadership in the post-pandemic era: Insights from high school leaders. *Interdisciplinary Journal of Education Research*, 6:1–14. https://doi.org/10.38140/ijer-2024.vol6.32

Organisation for Economic Co-operation and Development 2019. *Trends shaping education 2019*. Paris, France: OECD Publishing. https://doi.org/10.1787/trends_edu-2019-en

Organisation for Economic Co-operation and Development 2021. *Education at a glance 2021: OECD indicators*. Paris, France: OECD Publishing. https://doi.org/10.1787/b35a14e5-en

Palmgren PJ, Brodin U, Nilsson GN, Watson R & Stenfors T 2018. Investigating psychometric properties and dimensional structure of an educational environment measure (DREEM) using Mokken scale analysis – a pragmatic approach. *BMC Medical Education*, 18:235. https://doi.org/10.1186/s12909-018-1334-8

Peterson KD, Wahlquist C & Bone K 2000. Student surveys for school teacher evaluation. *Journal of Personnel Evaluation in Education*, 14:135–153. https://doi.org/10.1023/A:1008102519702

Saka AO & Onanuga PA 2019. Teacher effectiveness of some selected secondary schools' science, technology, engineering and mathematics subjects: Implication for sustainable development using science education. *Journal of Education in Black Sea Region*, 5(1):3–14. Available at https://files.eric.ed.gov/fulltext/ED602373.pdf. Accessed 16 February 2026.

Schleicher A 2018. *Valuing our teachers and raising their status: How communities can help. International summit on the teaching profession*. Paris, France: OECD Publishing. https://doi.org/10.1787/9789264292697-en

Schwab K 2017. *The fourth industrial revolution*. Geneva, Switzerland: World Economic Forum.

Schwab K & Davis N 2018. *Shaping the fourth industrial revolution*. Geneva, Switzerland: World Economic Forum.

Sijtsma K & Molenaar IW 2002. *Introduction to nonparametric item response theory* (Vol. 5). London, England: Sage.

Sijtsma K & Van der Ark LA 2017. A tutorial on how to do a Mokken scale analysis on your test and questionnaire data. *British Journal of Mathematical and Statistical Psychology*, 70(1):137–158. https://doi.org/10.1111/bmsp.12078

Straat JH, Van der Ark LA & Sijtsma K 2013. Comparing optimization algorithms for item selection in Mokken scale analysis. *Journal of Classification*, 30:75–99. https://doi.org/10.1007/s00357-013-9122-y

Stronge JH 2018. *Qualities of effective teachers* (3rd ed). Alexandria, VA: ASCD.

Tabatabaee-Yazdi M, Samir A, Bakhtshirin S & Tabatabaeiyazdi SM 2024. EFL teachers' attitudes towards mobile teaching affordances: A Mokken scale analysis. *Journal of Modern Languages*, 34(1):123–152. https://doi.org/10.22452/jml.vol34no1.7

Tavakol M & Dennick R 2011. Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2:53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Teo T, Unwin S, Scherer R & Gardiner V 2021. Initial teacher training for twenty-first century skills in the Fourth Industrial Revolution (IR 4.0): A scoping review. *Computers & Education*, 170:104223. https://doi.org/10.1016/j.compedu.2021.104223

United Nations Educational, Scientific and Cultural Organization 2021. *Reimagining our futures together: A new social contract for education*. Paris, France: Author. Available at https://unevoc.unesco.org/pub/futures_of_education_report_eng.pdf. Accessed 28 February 2026.

University of Johannesburg 2021. *Research ethics committee (REC) charter*. Available at https://www.uj.ac.za/research/research-ethics/. Accessed 9 May 2025.

Van der Ark LA 2007. Mokken scale analysis in R. *Journal of Statistical Software*, 20(11):1–19. https://doi.org/10.18637/jss.v020.i11

Van Schuur WH 2003. Mokken scale analysis: Between the Guttman scale and parametric Item Response Theory. *Political Analysis*, 11(2):139–163. https://doi.org/10.1093/pan/mpg002

Wind SA 2017. An instructional module on Mokken scale analysis. *Educational Measurement: Issues and Practice*, 36(2):50–66. https://doi.org/10.1111/emip.12153

World Medical Association 2013. World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA*, 310(20):2191–2194. https://doi.org/10.1001/jama.2013.281053

Worthington RL & Whittaker TA 2006. Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist*, 34(6):806–838. https://doi.org/10.1177/0011000006288127

Xu M, David JM & Kim SH 2018. The fourth industrial revolution: Opportunities and challenges. *International Journal of Financial Research*, 9(2):90–95. https://doi.org/10.5430/ijfr.v9n2p90